



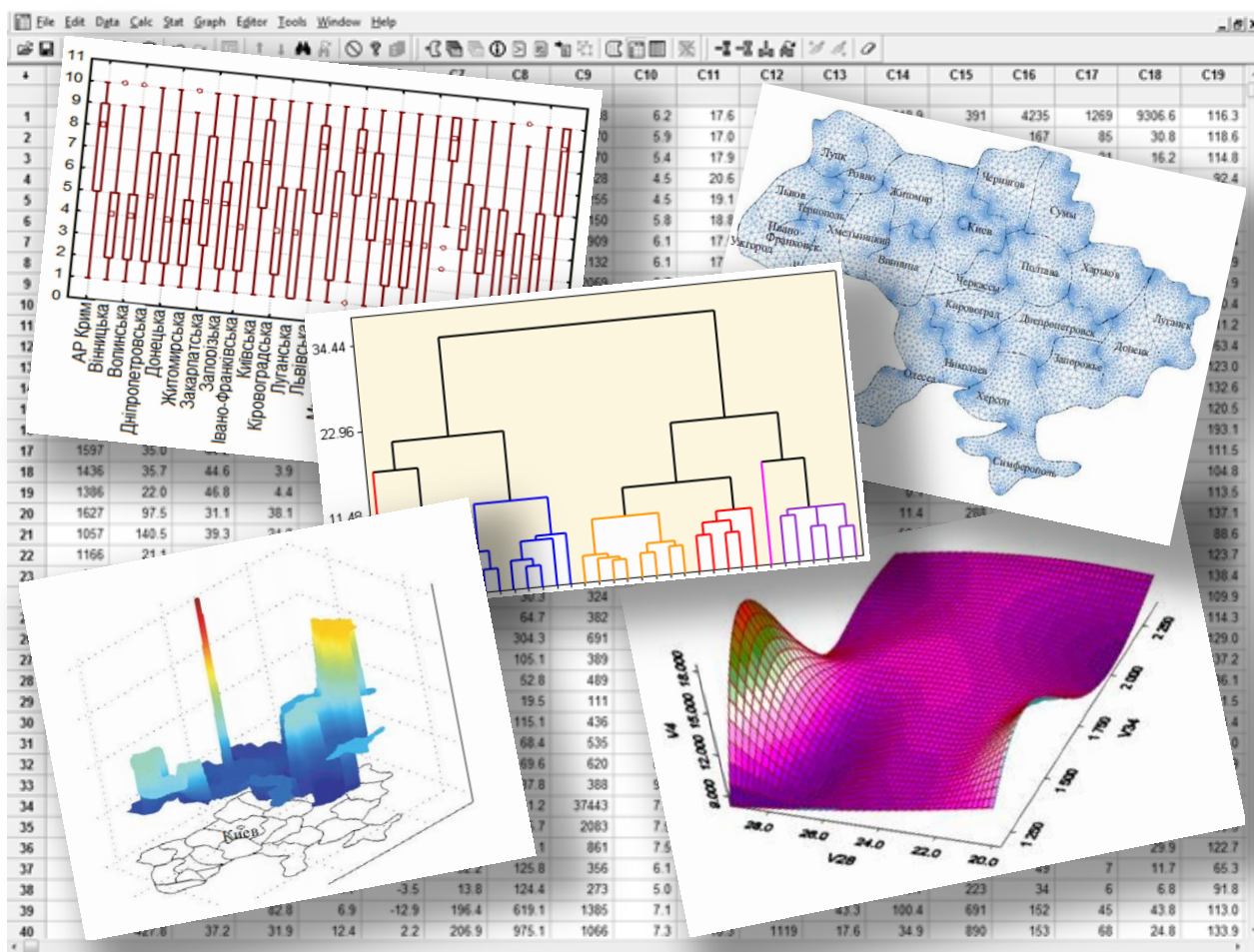
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И  
НАУКИ УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ  
УНИВЕРСИТЕТ имени В. Н. КАРАЗИНА



# СТАТИСТИЧЕСКИЕ МЕТОДЫ И ОБРАБОТКА ГЕОИНФОРМАЦИИ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ



Харьков – 2013

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ  
имени В. Н. КАРАЗИНА

# **СТАТИСТИЧЕСКИЕ МЕТОДЫ И ОБРАБОТКА ГЕОИНФОРМАЦИИ**

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

Харьков – 2013

УДК 91:004(075.8)  
ББК 26.8ф1я73  
Н50

**Рецензенты:**

- Голиков А. П. доктор географических наук, профессор,  
профессор кафедры международных экономических отношений  
факультета международных экономических отношений и  
туристического бизнеса  
Харьковского национального университета  
имени В. Н. Каразина
- Дейнека А. Г. доктор экономических наук, профессор,  
заведующий кафедрой менеджмента на транспорте  
факультета экономики транспорта  
Украинской государственной академии железнодорожного  
транспорта

**Немец К. А., Сегиды Е. Ю. Статистические методы и обработка геоинформации:** учебно-методическое пособие. – Харьков, Харьковский национальный университет имени В. Н. Каразина, 2013. – 60 с.

*Утверждено к печати решением Научно-методического совета  
Харьковского национального университета имени В. Н. Каразина  
(протокол № 1 от 16.10.2013 г.)*

Учебно-методическое пособие разработано в соответствии с программой курса «Статистические методы и обработка геоинформации» как одного из важных в общем цикле общественно-географической и общеэкономической подготовки специалистов.

Учебно-методическое пособие рассчитано на студентов дневного и заочного отделений геолого-географического факультета, обучающихся по специальности «Экономическая и социальная география». Пособие содержит общие сведения о курсе, тематический план (структуру) курса, учебную программу, список рекомендуемой литературы, перечень тем практическо-семинарских занятий, основные положения и методические указания к их выполнению, темы индивидуальных исследовательских заданий, вопросы к текущему и итоговому модульным контролям, систему оценки знаний студентов, а также краткий глоссарий основных понятий курса.

©Харьковский национальный университет  
имени В. Н. Каразина, 2013  
©Немец К. А., Сегиды Е. Ю., 2013

**СОДЕРЖАНИЕ**

Общие положения курса.....	4
Структура и рабочий план курса.....	6
Программа учебной дисциплины.....	7
Рекомендуемая литература.....	10
Темы и содержание практически-семинарских занятий.....	12
– Практическо-семинарское занятие № 1.....	12
– Практическо-семинарское занятие № 2.....	13
– Практическо-семинарское занятие № 3.....	18
– Практическо-семинарское занятие № 4.....	21
Вопросы к итоговому модульному контролю.....	36
Пример итогового модульного контроля.....	39
Оценка результатов учебных достижений студентов.....	41
Краткий глоссарий.....	42

## ОБЩИЕ ПОЛОЖЕНИЯ КУРСА

Одновременно со стремительным развитием науки, интеграцией отдельных ее отраслей, активно развивается методология общественной географии. Новый социальный запрос формирует необходимость комплексных исследований, привел к внедрению в общественную географию математико-статистических методов исследований. Для подготовки специалистов в области социальной и экономической географии в ВУЗах преподается курс «Статистические методы и обработка геоинформации», который является специальным в подготовке бакалавров географии и магистров (специалистов) социально-экономической географии. Он сочетает общенаучные подходы (математико-статистический, математический) с методологическими основами общественной географии, дает возможность создать у студента компетентность относительно обработки статистических данных, построения статистических моделей и корректной обработки геоинформационных данных. Курс «Статистические методы и обработка геоинформации» излагается параллельно с курсом «Теория и методология географической науки», что позволяет студентам понять роль и место статистических методов в географической методологии. В то же время, курс предшествует многим другим спецкурсам для специальности «Экономическая и социальная география», что создает благоприятные возможности для применения статистических методов в выполнении самостоятельных заданий из этих курсов.

**Цель курса** – формирование у студентов компетентности по применению статистических методов, математического моделирования и компьютерных технологий в общественно-географических исследованиях.

**Предметом курса** является изучение теоретических основ теории вероятностей и математической статистики как науки об исследовании случайных процессов и обработку случайных величин, а также о конкретных методах статистического моделирования общественно-географического процесса.

**Задания курса:**

- сформировать у студентов современную методологию использования математических методов, моделей и компьютерных технологий при исследовании социально-географических систем;
- дать студентам знания и понятия по основным методам и подходам в математической обработке общественно-географической информации;
- сформировать у студентов понятия о математических методах и моделирования при решении общественно-географических задач;

- сформировать у студентов компетентность по использованию компьютерных технологий и математических моделей в общественно-географических исследованиях.

В результате изучения данного курса:

**Студенты должны знать** понятия: наблюдение, описание, измерение, методология, гносеология, теоретическое знание, эмпирическое знание, парадигмы науки, практика, прогноз, познание, модель, математическая модель, геосистемная, синергетическая парадигма, информационно-синергетическая парадигма, идеализация, формализация, абстрагирование, синтез понятий, вероятность событий, закон больших чисел, статистическая устойчивость, системный анализ, выборочный метод, причинно-следственные связи, локальные процессы, интегральные процессы, плохо организованные системы, хорошо организованные системы, детерминированные процессы, случайные процессы, детерминированные модели, вероятностные модели, частота, распределение случайной величины, гистограммы, полигоны частот, центр распределения, математическое ожидание, дисперсия, стандарт, асимметрия, эксцесс, нормальное распределение, логарифмически нормальное распределение, линия регрессии, коэффициенты регрессии, ковариация, коэффициент корреляции, множественный и частичный коэффициент корреляции, графы, дендрограммы, распознавание образов, собственная область, гиперповерхность, обобщенное расстояние, специфичность и всеобщность переменных, свертка информации, поле пространственной переменной, закономерная и случайная изменчивость, тренд-анализ, фон, аномалия, генерализация фона, алгебраические и тригонометрические полиномы.

**Студенты должны уметь:** планировать общественно-географические исследования в зависимости от системы целей, возможностей исполнительной системы, конкретных условий, раскладывать сложные процессы на простые, описать причинно-следственные связи, определить вероятностный характер процессов, различать локальные и интегральные процессы, оценивать вероятность события, описать закон распределения случайной величины, применять методы статистического анализа и графически отображать результаты статистического исследования с помощью компьютера, использовать стандартные пакеты прикладных программ статистического анализа для решения конкретных прикладных задач общественной географии, использовать пространственные переменные для описания и анализа географических полей.

## СТРУКТУРА И РАБОЧИЙ ПЛАН КУРСА

Названия модулей и тем	Количество часов											
	Дневная форма						Заочная форма					
	Всего	в том числе					Всего	В том числе				
		л	п	лб	инд	ср		л	п	лб	инд	ср
1	2	3	4	5	6	7	8	9	10	11	12	13
<b>Модуль 1. Особенности использования математических методов (моделирование) в общественной географии</b>												
<b>Тема 1.</b> Место математической статистики в географической методологии	26	4	2			20	18	2				16
<b>Тема 2.</b> Особенности природных и общественных процессов, как объектов статистического моделирования	26	4	2			20	18	2				16
<i>Вместе по модулю 1</i>	52	8	4			40	40	4				36
<b>Модуль 2. Статистическое моделирование</b>												
<b>Тема 1.</b> Одномерные статистические модели и их особенности	19	4	2			13	14	2				12
<b>Тема 2.</b> Двумерные статистические модели и особенности их применения	22	4	2			16	18	2				16
<b>Тема 3.</b> Многомерные статистические модели	22	4	2			16	16	2	2			12
<b>Тема 4.</b> Моделирование пространственных переменных	29	6	3			20	20	2	2			16
<i>Вместе по модулю 2</i>	92	18	9			65	68	8	4			56
<b>Всего часов</b>	<b>144</b>	<b>26</b>	<b>13</b>			<b>66</b>	<b>108</b>	<b>12</b>	<b>4</b>			<b>92</b>

## **ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**

### **Модуль 1. Особенности использования математических методов (моделирования) в общественной географии**

#### **Тема 1. Место математической статистики в географической методологии**

*Лекция 1. Основные понятия и история использования статистических методов в общественной географии.* Объект и цель курса. Определение математических методов и моделирования. Понятие «модель». Классификации моделей: натурные, аналоговые, математические. История развития и использования моделей в естествознании и общественных науках.

*Лекция 2. Моделирование как метод познания, в общественной географии.* Три уровня математизации общественной географии: уровень параметризации географических явлений и объектов, уровень эмпирических моделей, уровень теоретических моделей. Место методов идеализации, формализации, математических методов, методов моделирования в методологии общественно-географического исследования. Противоречия системного и синергетического подходов. Перспективы развития моделирования в общественной географии.

#### **Тема 2. Особенности природных и общественных процессов, как объектов статистического моделирования**

*Лекция 3. Специфика развития природных и социальных объектов.* Специфика возникновения, функционирования и развития природных и социальных объектов. Многофакторность общественно-географических процессов. Сложность интерпретации причинно-следственных связей в социogeосистемах.

*Лекция 4. Случайный характер организации географической среды.* Организация географической среды. Локальные и интегральные процессы. Понятие о вероятности. Принципиально случайный характер общественно-географических процессов. Особенности общественно-географической информации.

### **Модуль 2. Статистическое моделирование**

#### **Тема 1. Одномерные статистические модели и их особенности**

*Лекция 5. Сущность одномерных статистических моделей.* Одномерные статистические модели. Понятие о случайной величине, дискретные и



непрерывные случайные величины. Числовые характеристики случайных величин: частота, функция распределения. Оценки центра распределения: математическое ожидание, медиана, мода. Оценки разброса случайной величины: центральные моменты второго, третьего и четвертого порядков. Основные законы одномерного распределения.

*Лекция 6. Оценка параметров общественно-географических объектов.* Выборочный метод, требования к выборочным совокупностям. Понятие о надежной вероятности. Точечная и интервальная оценка статистик. Требования к оценкам параметров общественно-географических объектов. Понятие о проверке статистических гипотез. Условия применения одномерных статистических моделей.

## **Тема 2. Двумерные статистические модели и особенности их применения**

*Лекция 7. Двумерные статистические модели.* Двумерные статистические модели. Понятие о двумерной случайной величине. Условия формирования двумерных случайных величин в общественно-географических исследованиях. Особенности формирования выборочных совокупностей. Особенности интерпретации связей компонентов двумерных случайных величин. Понятие о форме зависимости и силе связи двух случайных величин. Функциональные и статистические зависимости.

*Лекция 8. Регрессионно-корреляционный анализ и его применение.* Двумерный регрессионный анализ, понятие об условном распределении, линейные, нелинейные уравнения регрессии, вычисления регрессионных коэффициентов. Построение доверительного интервала для уравнений регрессии. Корреляционный анализ, понятие о коэффициенте корреляции, корреляционное отношение. Условия корректности корреляционного анализа. Нелинейные преобразования случайных величин для приведения к нормальному закону распределения. Особенности использования регрессионно-корреляционного анализа в общественно-географических исследованиях.

## **Тема 3. Многомерные статистические модели**

*Лекция 9. Многомерный статистический анализ.* Многомерные статистические модели. Понятие о многомерной случайной величине. Понятие о матрице исходных данных. Многомерный корреляционный анализ, парный, частичный, множественный коэффициент корреляции. Особенности использования многомерного корреляционного анализа. Многомерный

регрессионный анализ, определение значимости частичных коэффициентов регрессии. Особенности применения многомерного регрессионного анализа.

*Лекция 10. Методы многомерного статистического анализа, особенности их применения.* Понятие о многомерном признаковом пространстве, определение расстояния в многомерном пространстве. Использование теории графов для классификации общественно-географических объектов. Кластер-анализ, определение дистанционных коэффициентов, принципы образования кластеров. Задачи распознавания образов, понятия о собственной области объекта, решающей функции, решающее правило. Алгоритм распознавания образов. Дискриминантный анализ в задачах классификации. Факторный анализ, его разновидности, интерпретация результатов.

#### **Тема 4. Моделирование пространственных переменных**

*Лекция 11. Пространственные переменные как объект геоинформатики.* Понятие о пространственной переменной, признаки и свойства пространственных переменных. Различия моделей пространственных переменных от статистических моделей. Общественно-географические объекты как генераторы полей пространственных переменных. Изменчивость и анизотропность полей пространственных переменных. Закономерная и случайная составляющие наблюдаемой изменчивости.

*Лекция 12. Методы разделения фона и аномалий, методы расчета локальных показателей.* Понятие о фоне и аномалии. Понятие о тренде. Тренд-анализ как метод разделения фона и аномалии. Задачи тренд-анализа в общественной географии. Общая характеристика методов сглаживания случайных полей. Метод скользящего статистического окна. Принципы расчета локального среднего. Размер скользящего статистического окна как параметр генерализации карты аппроксимации случайного поля. Анализ остатков и выделение аномалий. Интерпретация результатов тренд - анализа.

*Лекция 13. Аппроксимация поверхности тренда единственной функцией координат пространства.* Обзор методов аппроксимации поверхности тренда единственной функцией координат пространства. Принципы оценки аппроксимации поверхности тренда. Модели алгебраических полиномов, выбор порядка полинома. Модели тригонометрических полиномов. Использование дифференциальных уравнений для аппроксимации поверхности тренда.

## РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Арманд А.Д. Информационные модели природных комплексов. – М.: Наука, 1975.
2. Арманд А.Д. Самоорганизация и саморегулирование географических систем. – М.: Наука, 1988.
3. Бендат Дж., Пирсон А. Прикладной анализ случайных данных. М., Мир, 1989.
4. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. М., Наука, 1983.
5. Бусленко Н.П. Моделирование сложных систем. М., Наука, 1976.
6. Вайдлих В. Социодинамика. Системный подход к математическому моделированию в социальных науках. – М.: Эдиториал УРСС, 2004.
7. Веников В.А. Теория подобия и моделирования (применительно к задачам электроэнергетики. – М.: Высшая школа, 1976.
8. Гродзинський М.Д. Стійкість геосистем до антропогенних навантажень. – Київ: Ліцей, 1995.
9. Гроп Д. Методы идентификации систем. Пер. с англ. В.А. Васильева, В.И. Лопатина. – М.: Мир, 1979.
10. Ивахненко А.Г., Мюллер И.А. Самоорганизация прогнозирующих моделей. Киев, Техника, 1985.
11. Исаченко А.Г. Оптимизация природной среды (географический аспект). – М., 1980.
12. Каждан А.Б., Гуськов О.И. Математические методы в геологии. М., Недра, 1990.
13. Капица С.П. Общая теория роста человечества. Сколько людей жило, живет и будет жить на Земле. – М.: Наука, 1999.
14. Математическое моделирование: проблемы и результаты. - М.: Наука, 2003.
15. Моделирование динамики геоэкосистем регионального уровня. Хомяков П.М. и др. – М.: Изд – во МГУ, 2000.
16. Моисеев Н.Н. Математические задачи системного анализа. М., Наука, 1981.
17. Немец К. А., Сегида Е. Ю. Статистические методы и обработка геоинформации: учебно-методическое пособие. – Харьков, Харьковский национальный университет имени В. Н. Каразина, 2013. – 80 с.
18. Немець К. А., Сегіда К. Ю. Статистичні методи і обробка геоінформації: навчально-методичний комплекс для самостійної роботи

студентів, які навчаються за напрямом підготовки «Географія», зі спеціальності «Економічна та соціальна географія». – Харків, 2012. – 64 с.

19. Немець Л.М., Немець К.А. До методики соціально-географічного асоціативного аналізу // Економічна та соціальна географія: міжвід. наук. зб. – К., 2003, - вип. 54. – с. 13 – 18.

20. Немець Л.М., Олійник Я.Б., Немець К.А. Просторова організація соціально – географічних процесів в Україні. - Київ – Харків: РВВ ХНУ, 2003.

21. Осауленко О.Г. Моделювання сталого розвитку соціально-економічних систем: Монографія. – К., 2001.

22. Пістун М.Д. Основи теорії суспільної географії. – К.: Вища школа, 1996.

23. Поздняков А.В., Черванев И.Г. Самоорганизация в развитии форм рельефа. – М.: Наука, 1990, 204 с.

24. Пригожин И., Стенгерс И. Порядок из хаоса. Новый диалог человека с природой. - М.: Эдиториал УРСС, 2001.

25. Пэнгл Р. Методы системного анализа окружающей среды. М., Мир, 1979.

26. Самарский А.А., Гулин А.В. Численные методы. М., Наука, 1989.

27. Сочава В.Б. Введение в учение о геосистемах. – Новосибирск: Наука, 1978.

28. Топчієв О.Г. Суспільно – географічні дослідження: методологія, методи, методики. Навч. посіб. – Одеса: Астропринт, 2005.

29. Хакен Г. Информация и самоорганизация. Макроскопический подход к сложным системам. – М.: Мир, 1991.

30. Харвей Дж. Модели развития пространственных систем в географии человека / Модели в географии. – М., 1971.

31. Шаблій О.І. Математичні методи в соціально-економічній географії: Навч. видання. – Львів: Світ, 1994.

32. Шаблій О.І. Суспільна географія: теорія, історія, українознавчі студії. - Львів: Львівський національний університет імені Івана Франка, 2001.

33. Шаблій О.І. Основи загальної суспільної географії. – Львів, 2003.

34. Шредингер Э. Пространственно-временная структура Вселенной. – Новокузнецк, ИО НФМИ, 2000.

35. Эткинс П. Порядок и беспорядок в природе. Пер. с англ. Ю.Г. Рудого. – М.: Мир, 1987.

## ТЕМЫ СЕМИНАРСКО-ПРАКТИЧЕСКИХ ЗАНЯТИЙ

№ п/п	Название темы	Количество часов
1	Осознание общественно-географической и математической сущности поставленной задачи, подготовка исходной матрицы данных.	2
2	Построение одномерных статистических моделей векторов переменных	3
3	Построение двумерных статистических моделей на матрице исходных данных	4
4	Кластер - анализ и факторный анализ исходных данных. Интерпретация полученных результатов	4
	<b>Итого</b>	<b>13</b>

Семинарско-практические занятия по курсу построены таким образом, чтобы способствовать последовательному усвоению теоретического материала по курсу, применению на практике математико-статистических методов. Достижение поставленной цели планируется путем последовательного выполнения следующих задач:

- Определение объекта и предмета исследования, формирование базы данных;
- Статистическая проверка выбранных показателей (одномерный анализ);
- Определение взаимосвязей и взаимозависимости между показателями (двумерный анализ);
- Определение территориальных и структурных особенностей общественно-географического процесса (кластер и фактор анализы);
- Интерпретация результатов, полученных путем математико-статистического исследования.

***ПРАКТИЧЕСКО-СЕМИНАРСКОЕ ЗАНЯТИЕ 1. Осознание общественно-географической и математической сущности поставленной задачи, подготовка исходной матрицы данных.***

1. Определение объекта и предмета исследования
2. Понятие «матрицы данных»
3. Требования к составлению матрицы данных
4. Отбор статистических показателей для анализа
5. Формирование исходной матрицы данных

Первой задачей практической работы является статистическое наблюдение – массовое или выборочное, планомерное, научно-организованное

наблюдение за массовыми явлениями, заключается в регистрации отобранных признаков у каждой единицы выборочной совокупности.

Первый этап – это ознакомление с теоретическим материалом курса, следующий – проведение выборочного сбора социально-экономических показателей по статистическим ежегодникам (Украины, Харьковской или любой другой области, исходя из выбранного предмета исследования). Статистический показатель – это обобщающая количественная характеристика свойств совокупности в целом или ее частей в частности применительно к конкретным условиям места и времени. Статистические показатели делятся на абсолютные, относительные и средние величины. Качественное содержание показателя определяется сущностью явления и находит свое отражение в названии. Количественную сторону представляют число и размерность (измеритель). Данное наблюдение является не сплошным, поскольку охватывает лишь часть общественно-экономических единиц, характеризующих выбранные районы.

Следующим шагом является построение информационной базы данных по статистической совокупности – это упорядоченное множество существующих элементов (однородных в определенном отношении), соединенных общими, присущими каждому элементу множества, свойствами, условиями и причинами существования и развития. Формирование качественной информационной базы – это фундамент статистического исследования, поскольку использование только объективной и достаточно полной информации на последующих этапах исследования позволяет получить правильные, обоснованные выводы о характере и закономерностях изучаемого процесса.

## ***ПРАКТИЧЕСКО-СЕМИНАРСКОЕ ЗАНЯТИЕ 2. Построение одномерных статистических моделей векторов переменных.***

Второй задачей практической работы является выявление зависимости динамики определенных показателей от времени, включает методы сглаживания и аналитического измерения. Сглаживание базируется на процедуре определения усредненной траектории развития показателей в прошлом и ее продолжении на будущее. Аналитическое измерение предполагает подбор математической функции, которая наилучшим образом отражает тенденцию динамики показателя во времени, и расчет на его основе прогнозных значений данного показателя. Выполняется с помощью одномерного анализа, осуществляется в программе MINITAB 14. Одномерный анализ относится к базовым, выполняется путем «Stat» → «Basic Statistics».

Одномерный анализ – содержит одну случайную величину (анализ одной случайной величины). В основе данного анализа является исследование свойств и характеристик рядов случайных величин, установление принадлежности этих рядов к определенному теоретическому распределению, определение сходства или различия данного ряда по сравнению с другими рядами.

Основными исходными составляющими одномерного анализа является понятие: простой статистический ряд (совокупность), упорядоченный статистический ряд (ранжированный), вариационный ряд, генеральная совокупность и выборка, объем выборки, частота и вероятность, закон распределения, критерии однородности и другие.

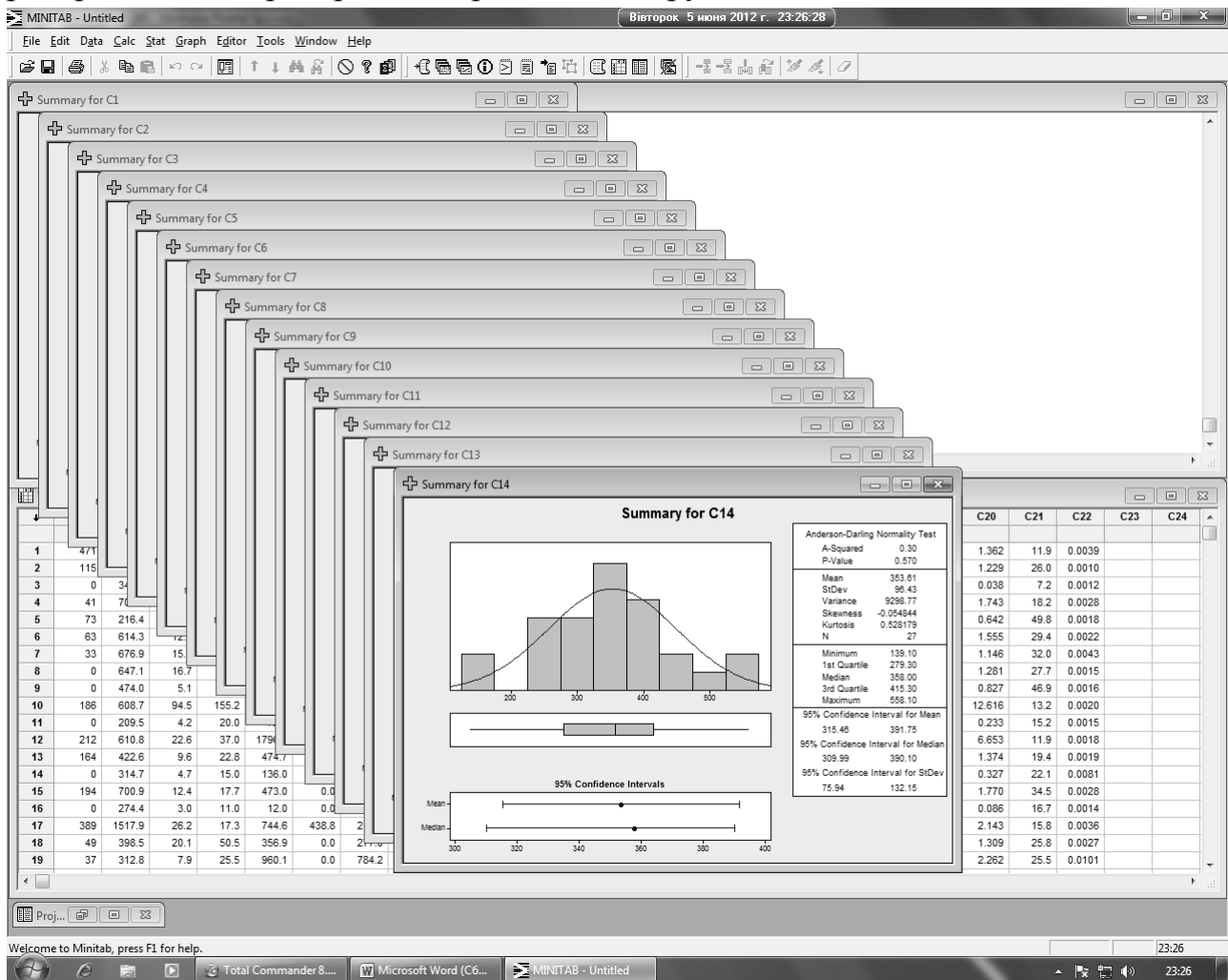


Рис. 1. Нормальное распределение статистического показателя

Совокупность случайных величин, полученных при одинаковых условиях проведения наблюдения, опытов, экспериментов называется простой статистической совокупностью или статистическим рядом. Статистический ряд – это первичная форма записи статистического материала в виде таблицы. Для наглядного представления исходного материала по данным статистического

ряда строится график изменений значений данной величины во времени или в пространстве или в хронологической последовательности.

Генеральная совокупность – это бесконечное или конечное число элементов или компонентов, состоящих из качественно однородных показателей. Любую часть генеральной совокупности, отобранную по определенным правилам, которая характеризует генеральную совокупность, называют статистической выборкой. Выборочную совокупность (выборку) создают обычно для облегчения обработки информации.

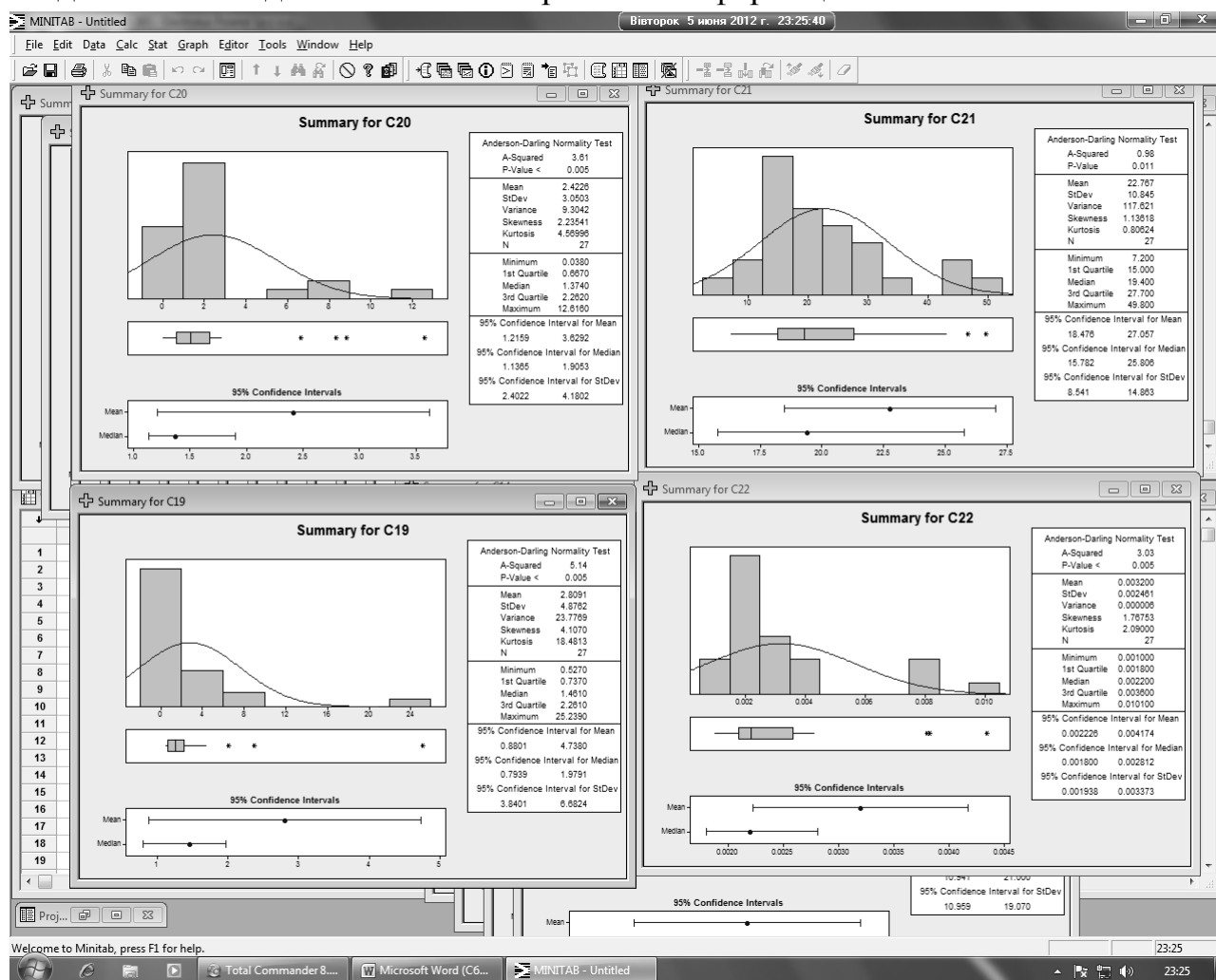


Рис. 2. Пример одномерного анализа статистических показателей

Одним из самых сложных и важных вопросов одномерного статистического анализа является определение количества наблюдений в исследованиях для получения надежного представления о характере изменчивости признака, который изучается, в генеральной совокупности. Обычно оптимальный объем выборки пропорционален степени изменчивости признака. Если признак сильно меняется, то количество измерений следует увеличить. Величину выборочной совокупности при выполнении



географических исследований можно определить двумя способами: по таблице достаточно больших чисел; расчетным способом. В обоих случаях количество наблюдений (численность или объем выборки) определяется, исходя из доверительной вероятности.

Систематизация и упорядочение данных, представляющих статистическую совокупность. Приведение их в определенную систему и характеристика этой системы. Всякая статистическая совокупность характеризуется объемом ряда и абсолютной частотой повторения одинакового признака. Частота – это число, которое показывает сколько раз встречается данный признак в изучаемой совокупности. Закон изменения частоты – это и есть закон распределения. Существует три способа представления закона распределения: табличный, графический и аналитический. Результаты обработки исходных данных, как правило, сначала всегда оформляются в виде статистических таблиц. Такая форма позволяет придать материалу удобство, компактность и рациональность.

Способ, который позволяет в наглядной форме получить представление о закономерностях распределения, называется графическим изображением вариационного ряда. Существует несколько способов графического изображения рядов распределения. При этом обычно используют данные таблицы эмпирического распределения. При графическом изображении рядов распределения на горизонтальной оси откладываются значения интервалов или наблюдений значения случайной величины, а на вертикальной оси – частоты. Для наглядного представления вариационного ряда чаще используют графические изображения в виде гистограммы, полигона, кумуляты, кривой концентрации Лоренса и другие. Гистограмма наглядно показывает распределение изучаемых величин, из-за чего подобный способ уже часто используется для иллюстрации особенностей статистического распределения. Вариационный ряд при этом изображается в виде столбиков, границы между которыми проходят по координатах, соответствующих границам между классами. При этом основание столбиков по ширине равно величине интервала признака, а высота – пропорциональна частоте отдельных классов. Рассмотрим основные статистические характеристики вариационных рядов. Одним из основных параметров статистического ряда является среднее значение признака или центр, относительно которого распределяются члены совокупности. Значения средних при изучении различного рода закономерностей географических явлений, процессов и объектов очень велико. Они позволяют: определить общую тенденцию развития явлений; оценивать значение отдельной величины путем сравнения ее со средней, определять

наличие связи между явлениями с помощью анализа средних двух или нескольких признаков по территориям или временным промежуткам.

При обработке данных географических исследований как важнейших характеристик вариационного ряда применяются различные средние значения. Средние величины принято разделять на простые и взвешенные. Если средние значения вычисляются по непосредственному перечню значений признака в каждой единице совокупности, то такие средние называются простыми (невзвешенными). Если средние вычисляются по вариационному ряду с учетом статистического веса каждого варианта, то их называют взвешенными.

Средние величины бывают разного рода. Если вид средней не указан, то подразумевается средняя арифметическая величина.

Средние величины чаще используют степенные и структурные (порядковые) средние. Степенные средние, в свою очередь, подразделяются на средние арифметические, средние гармонические, средние квадратичные, средние кубические и другие. Структурные средние – это мода, медиана, квартили, децили и др.

Медианой называется среднее (срединное) значение признака ранжированного вариационного ряда, то есть значение, равноудаленное от начала и конца, перестроенного в возрастающем или убывающем порядке.

Модой называется вероятная (что чаще встречается) в данном статистическом ряду величина. Мода является наибольшей ординатой кривой распределения при одновершинном распределении. В общем случае кривая распределения может иметь несколько вершин и, соответственно, она будет иметь несколько мод.

Считается, что нормальное распределение представляет собой одну из эмпирически проверенных истин и его положения рассматриваются как один из фундаментальных законов природы. Точная форма нормального распределения – «колоколообразная кривая» – определяется только двумя параметрами: средним и стандартным отклонением. Для нормального распределения характерно, что 68% всех значений находятся в пределах  $\pm 1$  стандартное отклонение от среднего, а диапазон  $\pm 2$  стандартных отклонения включает 95% значений. Об отклонении от нормального распределения свидетельствуют коэффициенты асимметрии и эксцесса. Коэффициент асимметрии показывает отклонение распределения от симметричного (нормальное распределение – абсолютно симметричное, соответственно коэффициент асимметрии равен нулю). Положительное значение коэффициента асимметрии свидетельствует о наличии распределения с «длинным правым хвостом», отрицательное – с «длинным левым хвостом». Коэффициент эксцесса показывает «остроту пика»

распределения (для нормального распределения коэффициент эксцесса также равен нулю). Если коэффициент эксцесса является положительным, пик является заостренным, если является отрицательным – пик закруглен.

**ПРАКТИЧЕСКО-СЕМИНАРСКОЕ ЗАНЯТИЕ 3.** *Построение двумерных статистических моделей на матрице исходных данных.* Третьей задачей практической работы было определение пар показателей, связанных между собой, определение количественной оценки силы связи между признаками. Двумерный анализ – содержит две случайные величины, образует двумерную модель и анализирует две величины. В ходе практической работы изучаются такие двумерные модели как: корреляционный и ковариационный анализы, выполняемой в программе MINITAB 14. путем «Stat» → «Basic Statistics» → «Correlation».

Корреляционный анализ – метод обработки статистических данных, заключающийся в изучении тесноты связи между переменными, при этом сравниваются коэффициенты корреляции между одной парой или множеством пар признаков для установления между ними статистической взаимосвязи, позволяет выявить наиболее связанные переменные и построить поверхности их зависимости.

Цель корреляционного анализа – обеспечить получение некоторой информации об одной переменной с помощью другой переменной. В случаях, когда возможно достижение цели, говорят, что переменные коррелируют. В самом общем виде восприятия гипотезы о наличии корреляции означает, что изменение значения переменной А произойдет одновременно с пропорциональным изменением значения В. Мерой зависимости между экспериментальными наборами данных являются числа – коэффициенты связи.

Главные задачи корреляционного анализа:

- 1) оценка по выборочным данным коэффициентов корреляции;
- 2) проверка значимости выборочных коэффициентов корреляции или корреляционное отношение;
- 3) оценка близости обнаруженной связи к линейной;
- 4) построение доверительного интервала для коэффициентов корреляции.

Определение силы и направления взаимосвязи между переменными является одной из важных проблем анализа данных. В общем случае для этого применяют понятие корреляции. Корреляция является зависимостью двух случайных величин. При этом, изменение одной или нескольких этих величин приводит к систематическому изменению другой или других величин.

Математической мерой корреляции двух случайных величин служит коэффициент корреляции.

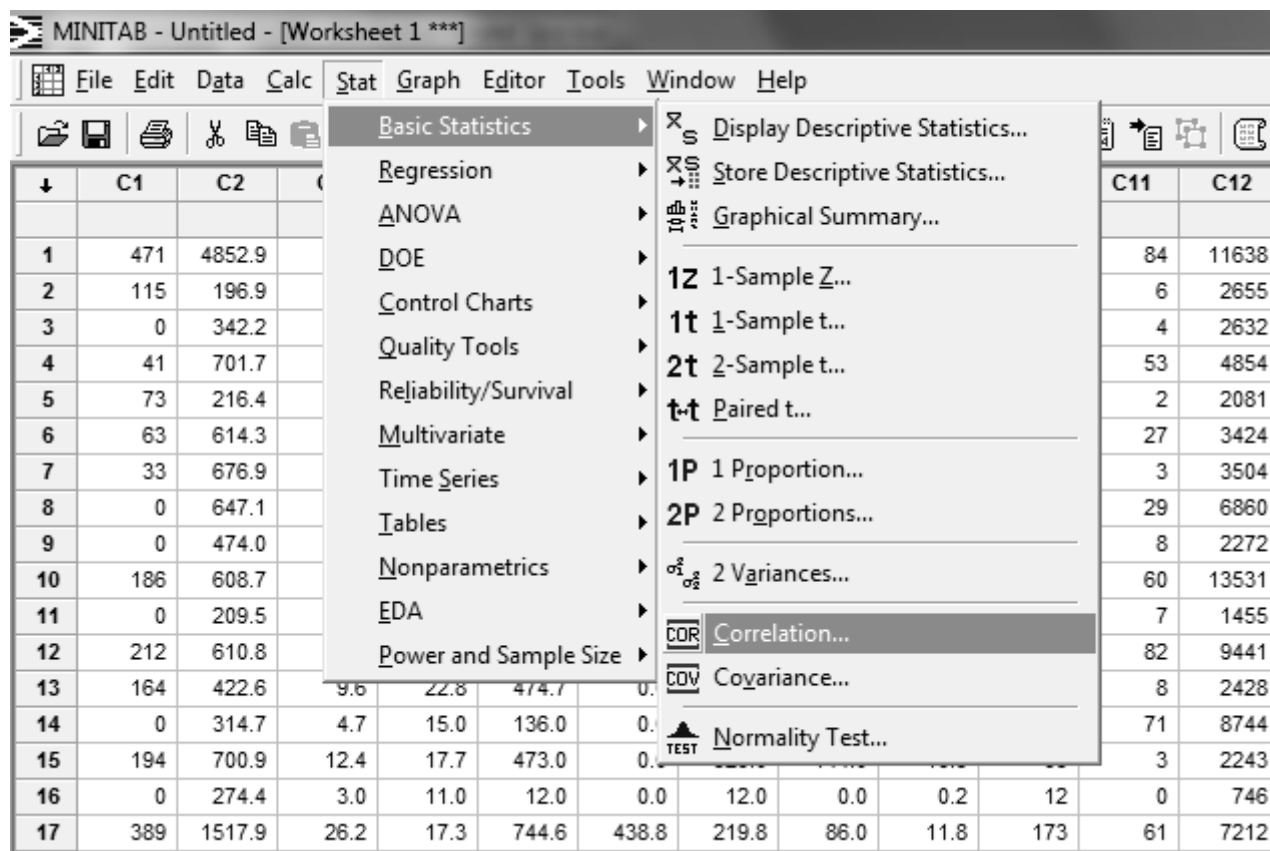


Рис. 3. Алгоритм выбора корреляционного анализа в программе MINITAB 1.

Коэффициент корреляции принимает значения в пределах от -1 до +1. Знак «-» означает обратную зависимость, «+» – прямую. Если коэффициент равен нулю, то линейная связь между динамическими рядами является отсутствующей, а если единице – имеется функциональная зависимость. Довольно часто ошибочно считается, что значение коэффициента корреляции меньше 0.2 свидетельствует об отсутствии связи. На самом деле он свидетельствует о наличии очень слабой связи, но связь между показателями имеется.

Коэффициент

корреляции

$> \pm 0.91$

$\pm 0.71-0.90$

$\pm 0.51-0.70$

$\pm 0.21-0.50$

$< \pm 0.20$

Теснота связи

Очень сильная

Сильная

Относительная

Слабая

Очень слабая

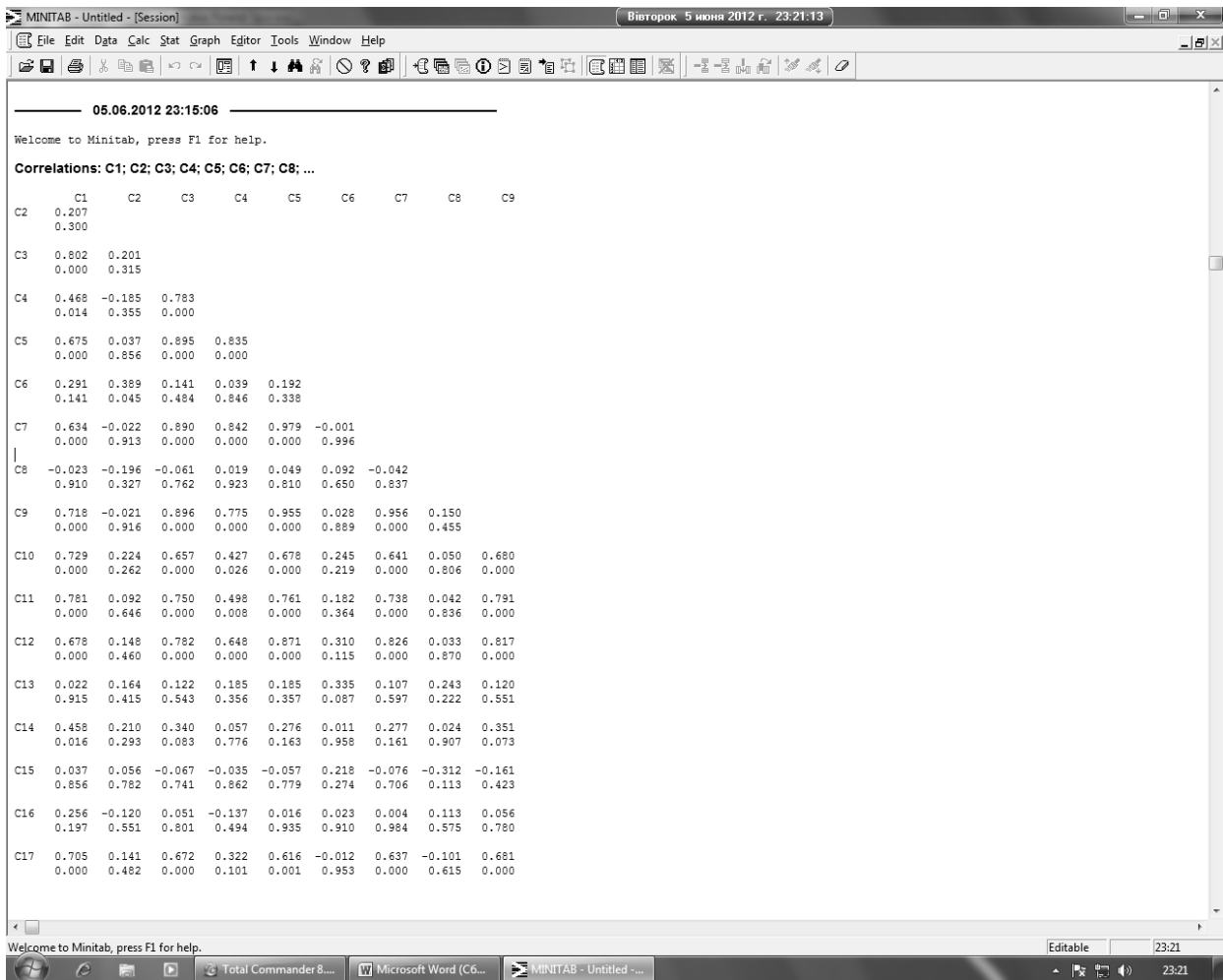


Рис. 4. Корреляционная матрица

Коэффициент корреляции, а в общем случае корреляционная функция, позволяет установить степень взаимосвязи между переменными. Корреляция может быть линейной или нелинейной в зависимости от типа зависимости, которая фактически существует между переменными. Довольно часто на практике рассматривают только линейную корреляцию (взаимосвязь), но более глубокий анализ требует использования для исследования процессов нелинейных зависимостей. Сложную нелинейную зависимость можно упростить, но знать о ее существовании необходимо для того, чтобы построить адекватную модель процесса.

В случае максимальной тесноты связи между показателями на диаграмме рассеяния их зависимость будет представлена прямой линией. Иногда встречается такое явление, как псевдокорреляция, то есть корреляция, обусловленная влиянием других показателей, оставшихся вне поля зрения исследователя. Значимость коэффициента корреляции зависит от длины динамических рядов. В больших динамических рядах даже слабые зависимости будут значимыми, в то время, как в незначительных — даже очень сильные зависимости не являются статистически надежными. Поэтому говорят о

надежности корреляционной зависимости, связанной с репрезентативностью исходных динамических рядов и свидетельствует о том, насколько вероятно, что выявленная зависимость, опять будет обнаружена при увеличении периода ретроспекции или экстраполяции на будущее. Надежность выявленных зависимостей оценивается с помощью стандартной статистической меры –  $p$ -уровня – статистического уровня значимости. Данный показатель находится в обратной зависимости к надежности результата: чем выше  $p$ -уровень, тем ниже статистическая надежность обнаруженной зависимости, и наоборот.

Ковариация – в теории вероятностей и математической статистике, числовая характеристика зависимости случайных величин. Сущность ковариации заключается в том, что она возникает вследствие неопределенности результата перемножения двух совокупностей чисел.

Построение ковариационной матрицы достаточно похоже на построение матрицы корреляции, выполняется в программе MINITAB 14 путем «Stat» → «Basic Statistics» → «Covariance». По сравнению с корреляционной связкой ковариационная матрица не указывает на тесноту связи между определенными группами факторов, она лишь указывает на результат перемножения двух совокупностей чисел.

#### ***ПРАКТИЧЕСКО-СЕМИНАРСКОЕ ЗАНЯТИЕ 4. Кластер-анализ и факторный анализ исходных данных. Интерпретация полученных результатов.***

Четвертым заданием практической работы будет определение территориальной дифференциации административно-территориальных единиц и выявление значимых факторов общественно-географического процесса.

Многомерным анализом называют совокупность различных методов, предназначенных для изучения многомерных явлений. В многомерном пространстве исследуемые объекты располагаются, как правило, не равномерно, а образуют определенные скопления. Эти скопления можно рассматривать как классы объектов. Причем в разных пространствах выделяются различные классы. В многомерном пространстве классификация более обоснована. Но это означает, что при безграничном возрастании числа признаков в той же пропорции растет точность классификации.

Существует много математических методов и приемов, которые на основе информации заложенной в матрице данных, позволяют объективно классифицировать экономико-географические объекты (в том числе регионализировать их). Такие методы и приемы объединены в многомерный анализ, рассматривается в узком и широком смыслах. В узком смысле – это

анализ матрицы данных, имеет два и более столбцов, в широком смысле – это анализ матрицы данных, когда количество столбцов не ограничено снизу. В этом случае многомерный анализ включает в себя и одномерный. Итак, одномерный анализ можно рассматривать как частный случай многомерного, а многомерный – как обобщение одномерного. Часто в процессе обработки матрицы данных приходится их сворачивать, то есть сводить многомерное пространство к  $n$ - $k$ -мерному, и даже к одномерному. В этом случае много признаков заменяются несколькими, но такими, каждая из которых синтезирует информацию соответствующей группы признаков или является наиболее репрезентативной.

*Кластерный анализ* предполагает группировку объектов по сходству определенных характеристик. Термин впервые введен Трайеном / Tryon /, 1939 г. Это один из методов классификации, предусматривающий разделение исходной совокупности объектов на кластеры (классы, группы). Кластер – это группа территориальных единиц (регионов), имеющих сходные тенденции или особенности развития. С математико-статистической точки зрения, каждый кластер должен иметь следующие свойства: плотность объектов в пределах кластера должен быть выше плотности вне его, возможность отделения от других кластеров и т.д. Сложность задач кластерного анализа заключается в том, что реальные общественно-географические объекты являются многомерными, т.е. описываются не одним, а определенной совокупностью параметров, поэтому объединение в группы осуществляется в пространстве многих измерений. Согласно критерию объединения регионов в кластеры есть минимум расстояния в пространстве показателей, которые их описывают. Отсюда – понятие расстояния между регионами в пространстве данных.

В ходе практической работы выполняется проведённый кластерный анализ, в программе MINITAB «Stat» → «Multivariate» → «Cluster Obrevations».

Существуют следующие группы методов кластерного анализа:

- 1) иерархические методы;
- 2) итеративные методы;
- 3) факторные методы;
- 4) методы сгущений;
- 5) методы, использующие теорию графов.

К распространенным в экономике относят иерархические и итеративные.

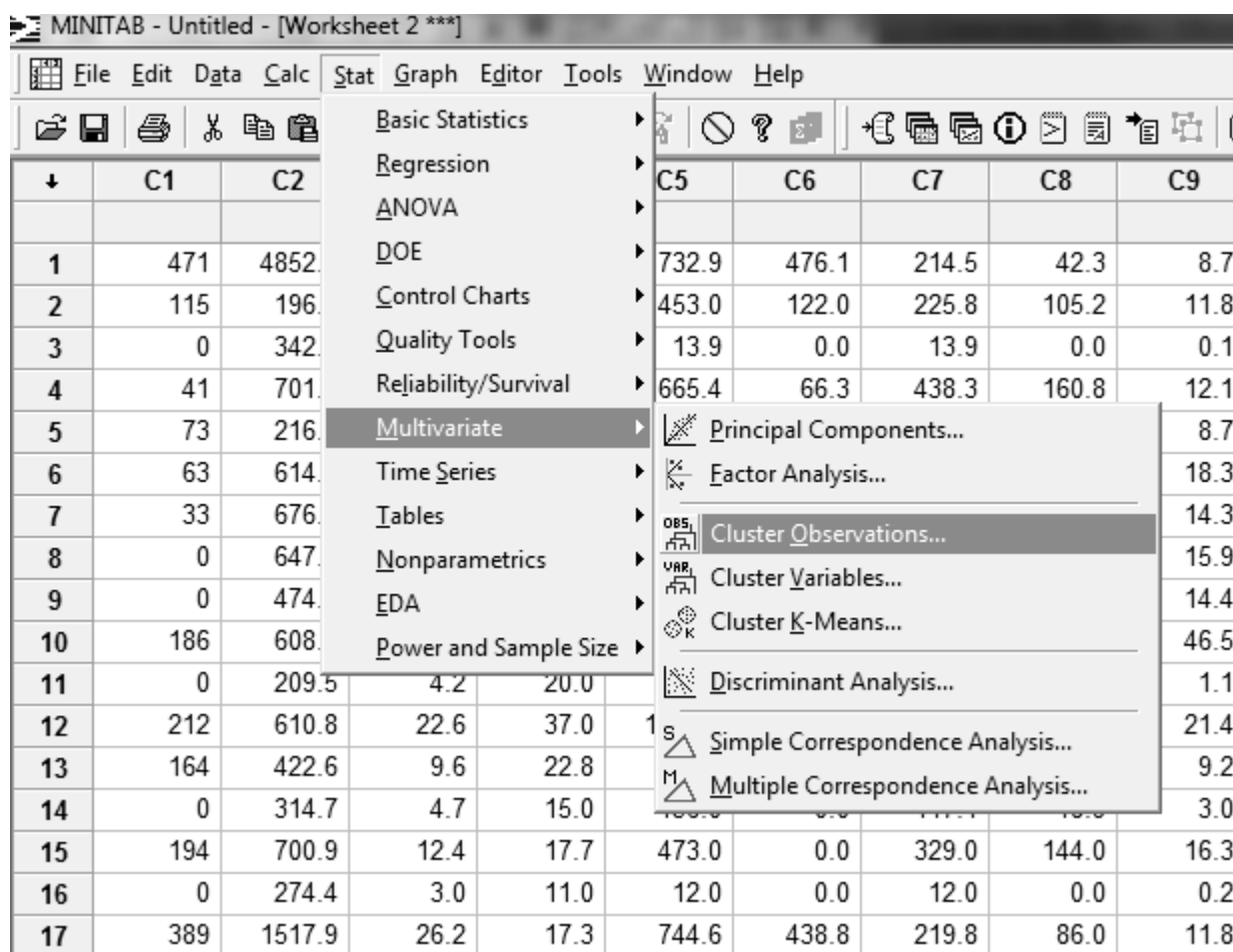


Рис. 5. Алгоритм выбора кластерного анализа в программе MINITAB 14

Для проведения классификации необходимо ввести понятие сходства объектов по наблюдаемым переменным. В каждый кластер (класс, таксон) должны попасть объекты, имеющие подобные характеристики.

Выбор расстояния между объектами является узловым моментом исследования, от него во многом зависит окончательный вариант разбиения объектов на классы при данном алгоритме разбиения.

Алгоритм иерархического агломеративного кластерного анализа можно представить в виде последовательности процедур:

- 1) нормируются исходные данные;
- 2) рассчитывается матрица расстояний или матрица мер сходства;
- 3) находится пара ближайших кластеров, по выбранному алгоритму объединяются эти два кластера. Новому кластеру присваивается меньший из номеров объединяемых кластеров;
- 4) процедуры 2, 3 и 4 повторяются до тех пор, пока все объекты не будут объединены в один кластер или до достижения заданного "порога" сходства.

На расчет расстояний между объектами существенное влияние имеет выбор единиц измерения показателей. Так, например, если брать показатель объемов производства продукции в миллионах гривен, разница между двумя



территориальными единицами по этой координатой будет одной, если же в тысячах гривен – в тысячу раз больше, что влияет на конечное группировки. Поэтому кластерный анализ предусматривает осуществление процедуры нормализации данных.

В кластерном анализе для количественной оценки сходства вводится понятие метрики. Сходство или различие между классифицированными объектами устанавливается в зависимости от метрического расстояния между ними. Если каждый объект описывается  $k$  признаками, то он может быть представлен как точка в  $k$ -мерном пространстве, и сходство с другими объектами будет определяться как соответствующее расстояние. В кластерном анализе используются различные меры расстояния между объектами.

Наиболее распространенными являются следующие виды расстояний:

- евклидовое расстояние (Euclidean distances), рассчитывается по теореме Пифагора: расстояния по каждой из координат вводятся в квадрат, а затем с их суммы определяется корень квадратный
- манхэттенское расстояние (расстояние городских кварталов, city-block/Manhattan/distances), рассчитывается как сумма разниц по каждой из координат;
- расстояние Чебышева (Chebychev distance metric), регионы определяют как "различные", если они различаются по какой-либо одной координате;
- процент несогласия (percent disagreement), используется, когда исходные данные не имеют количественного выражения.

Методы кластеризации делятся на две большие группы – агломеративные (от англ. Agglomerate – скопление) и дивизивные (от англ. Division – разделение).

Агломеративные методы предполагают последовательное объединение подобнейших регионов на основе рассчитанных расстояний между ними. Процедура кластеризации такая. На первом шаге, каждый регион образует отдельный кластер, далее в новый кластер объединяются два региона, степень сходства которых является наибольшей. На последнем шаге все регионы объединяются в один кластер.

Мера сходства кластеров и регионов определяется следующими способами:

- единичной связи (single linkage, «метод ближайшего соседа»): минимум наименьших расстояний до какого-либо одного региона в кластере;
- полной связи (complete linkage, «метод самых удаленных соседей»): минимум наибольших расстояний до какого-либо одного региона в кластере;
- «средней» связи: минимум среднеарифметического значения расстояний ко всем регионам в кластере;

- центроидный минимум расстояний к центрам тяжести кластеров. Центр тяжести кластера определяется как среднее по каждому параметру.

Результат кластеризации визуализируется в виде дендрограммы кластеризации (tree diagram, дерево объединения), на одной оси которой откладываются регионы, на второй – расстояния объединения (linkage distance).

Дивизионные методы предусматривают поэтапное (итерационное) разделение регионов на заданное количество кластеров. Распространенным среди дивизионных является метод  $k$ -средних (kmeans clustering), который решает задачу выделения заранее заданного количества кластеров, которые максимально отличаются, т.е. находятся на крупнейших расстояниях друг от друга.

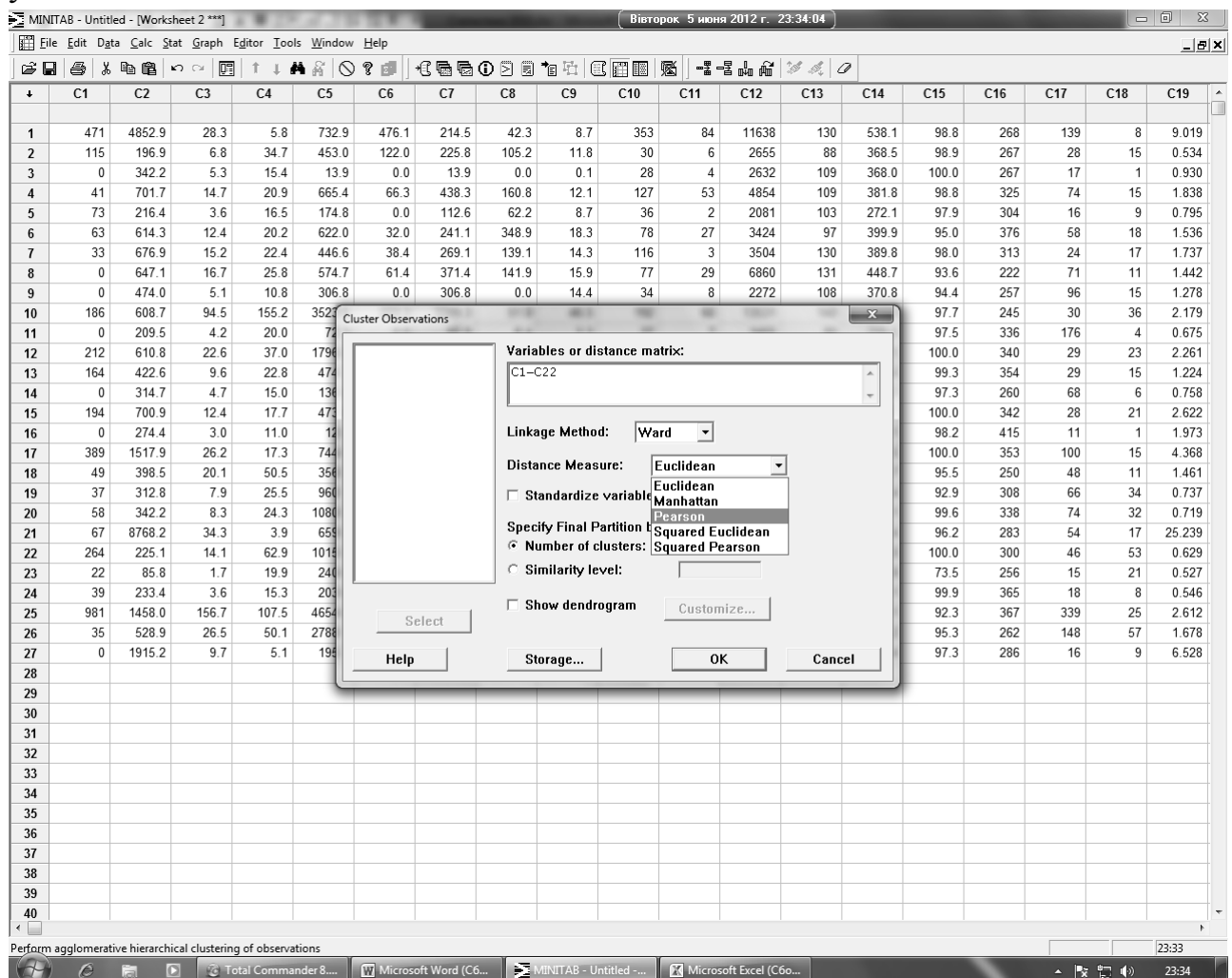


Рис. 6. Выбор расстояния при выполнении кластерного анализа

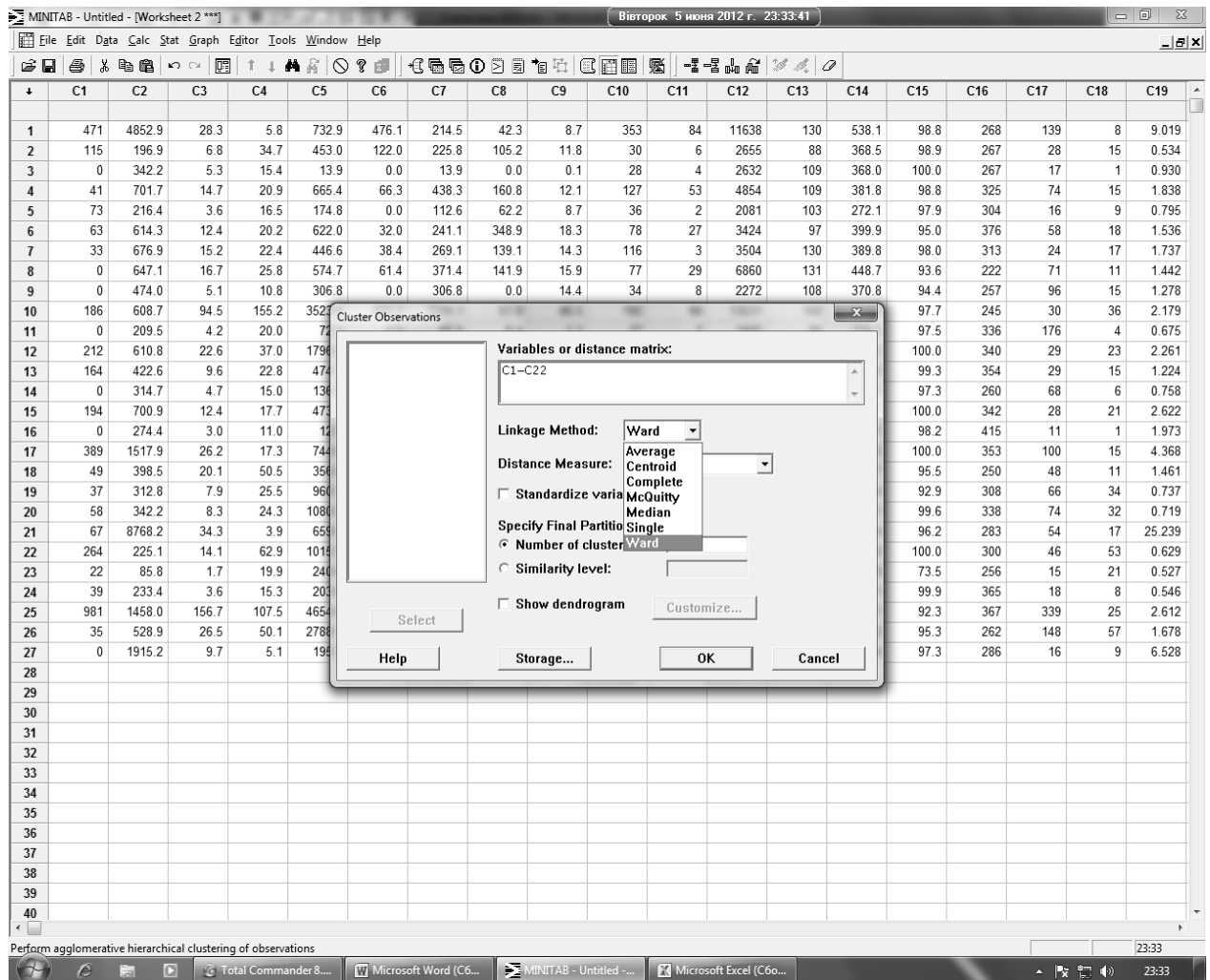


Рис. 7. Выбор метода при выполнении кластерного анализа

Процедура кластеризации такая. На первом шаге задается некоторое случайное разделение данных на заданное количество кластеров ( $k$ ), рассчитываются центры тяжести кластеров. Далее осуществляется перемещение регионов: каждый регион относится к тому кластеру, расстояние до центра тяжести которого минимальное, рассчитываются центры тяжести новых кластеров. Эта процедура повторяется, пока не будет найдена стабильная конфигурация, т.е. состав кластеров перестанет изменяться.

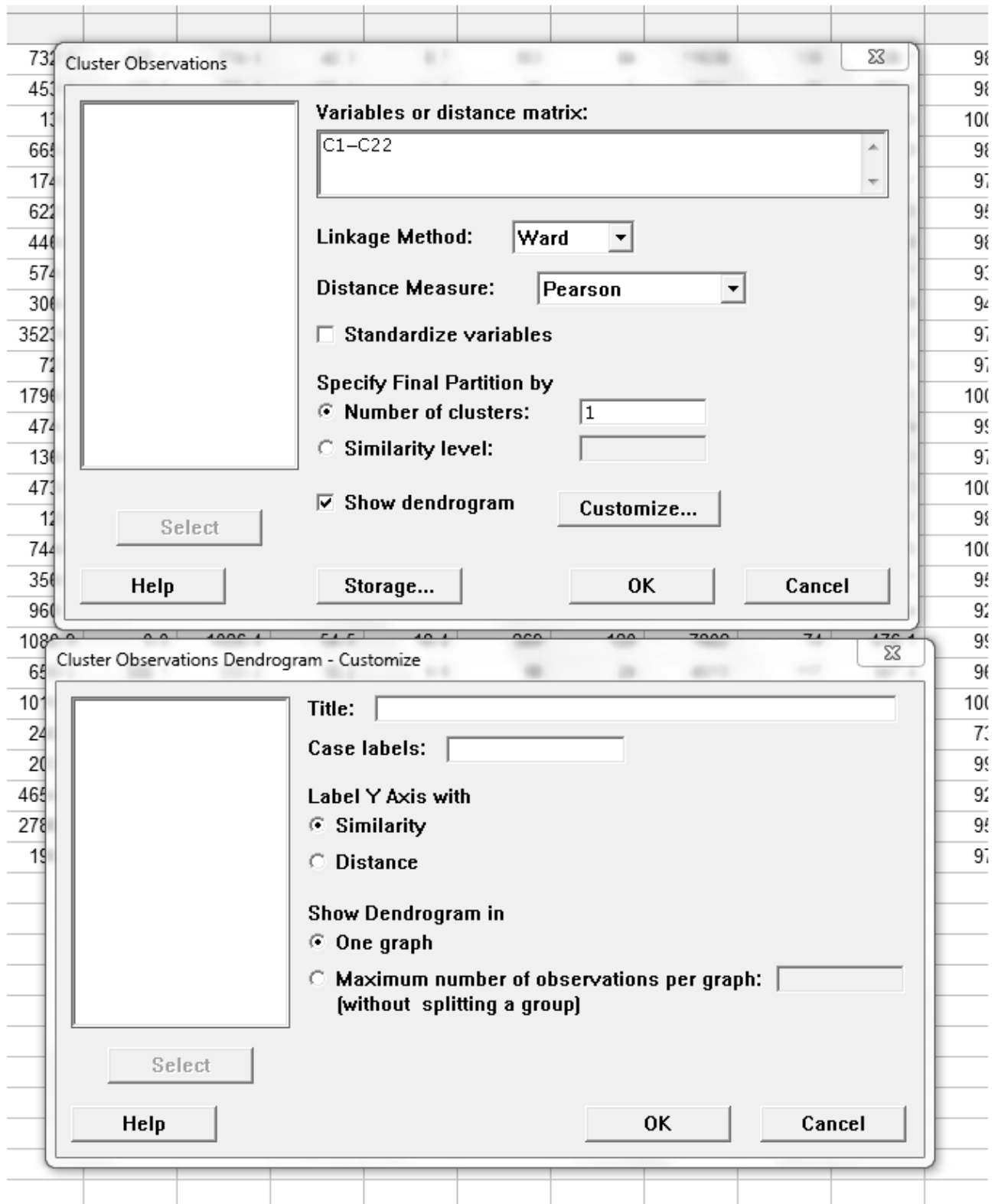


Рис. 8. Выбор основных настроек при выполнении кластерного анализа

Для визуализации результатов строятся графики, представляющие собой проекции любой пары показателей на плоскость, на которых точки, относящиеся к одному кластеру, окантовываются. Одним из самых важных и сложных вопросов при кластеризации является выбор оптимального количества

кластеров. Обычно согласно исходной гипотезы определяется начальное количество кластеров, а затем изменяя его, эмпирическим путем выбирают окончательный вариант кластеризации. Для того, чтобы оценить, насколько выделенные кластеры отличаются, рассчитывают средние значения базисных показателей для каждого кластера.

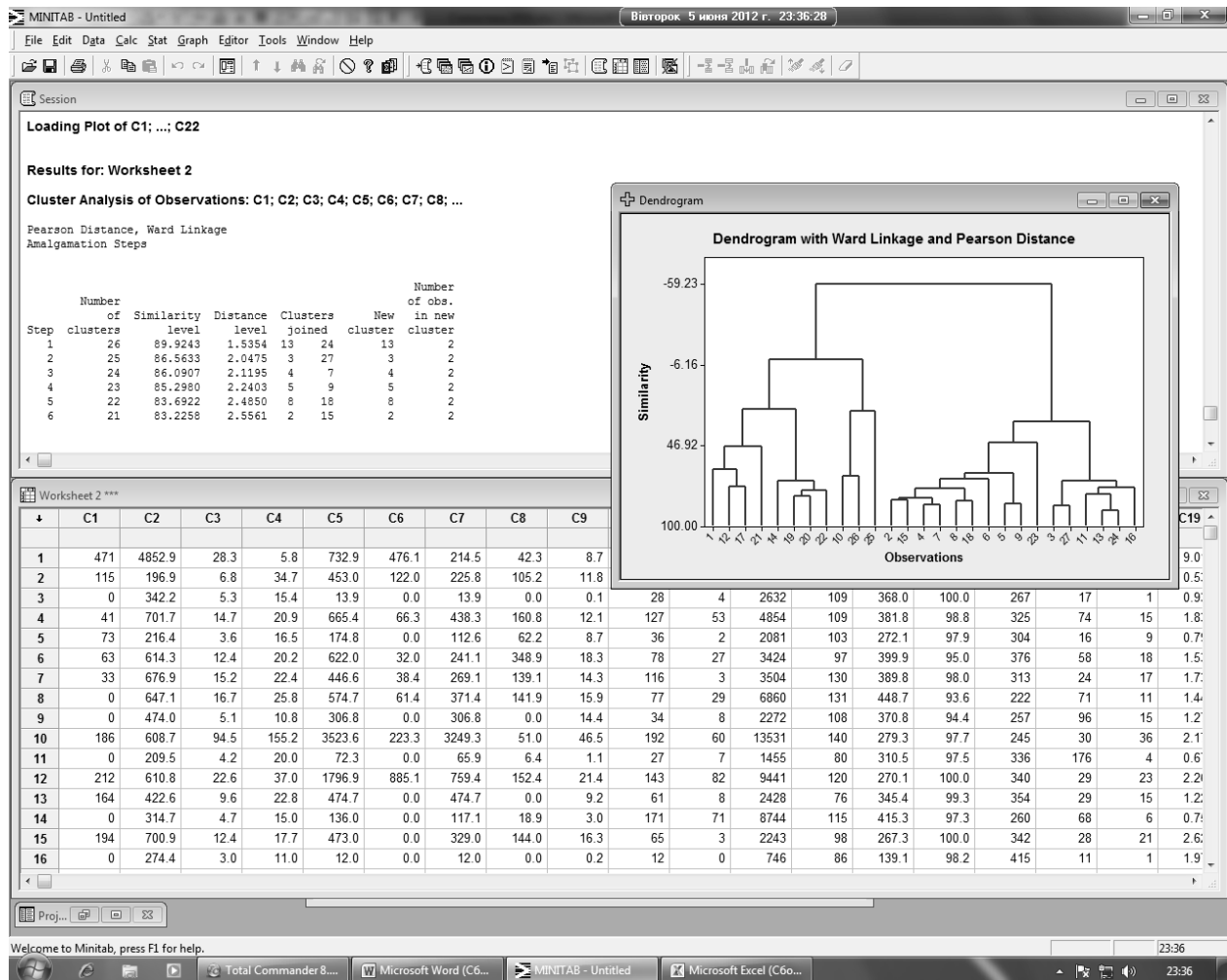


Рис. 9. Пример результата кластерного анализа

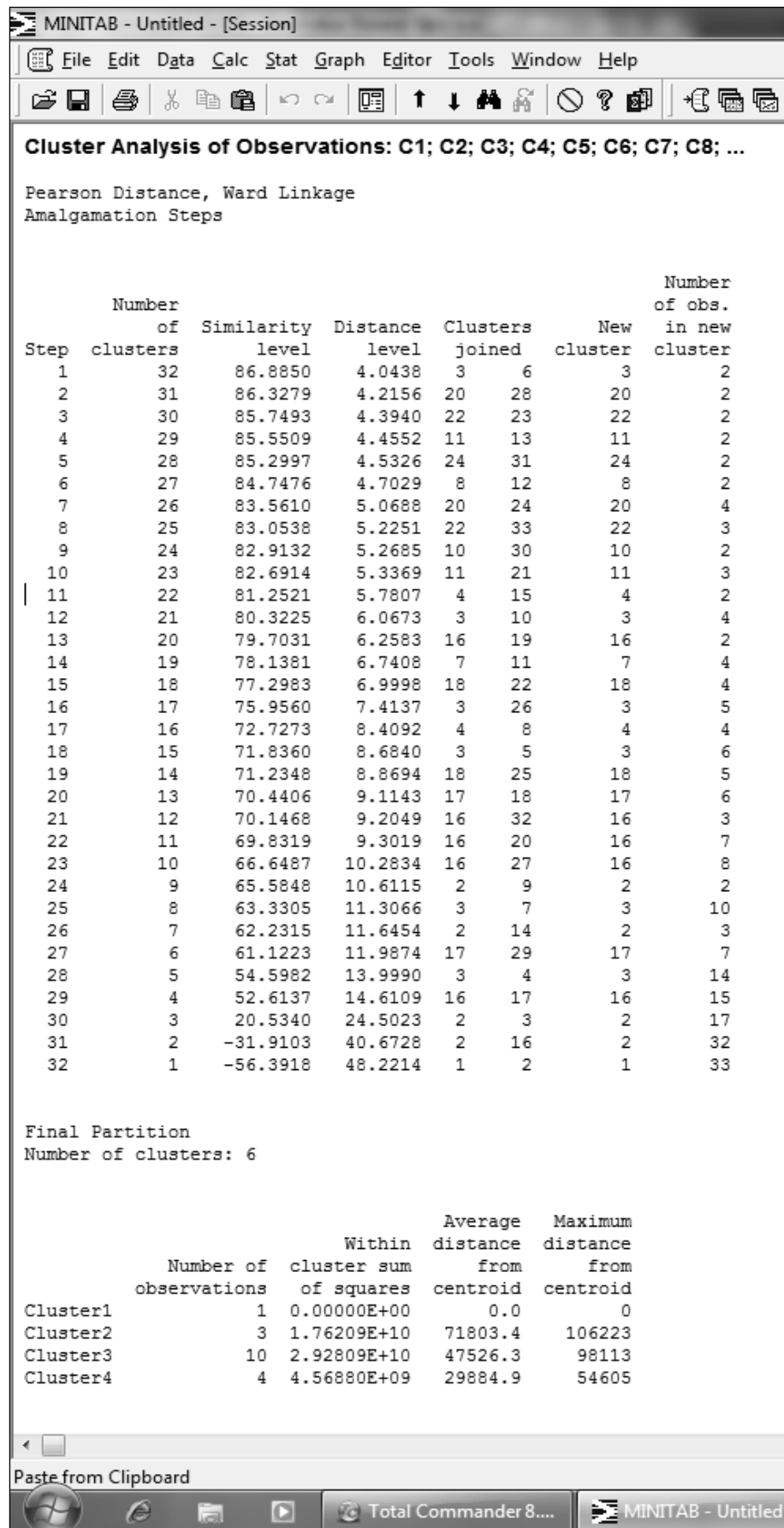


Рис. 10. Пример протокола кластерного анализа

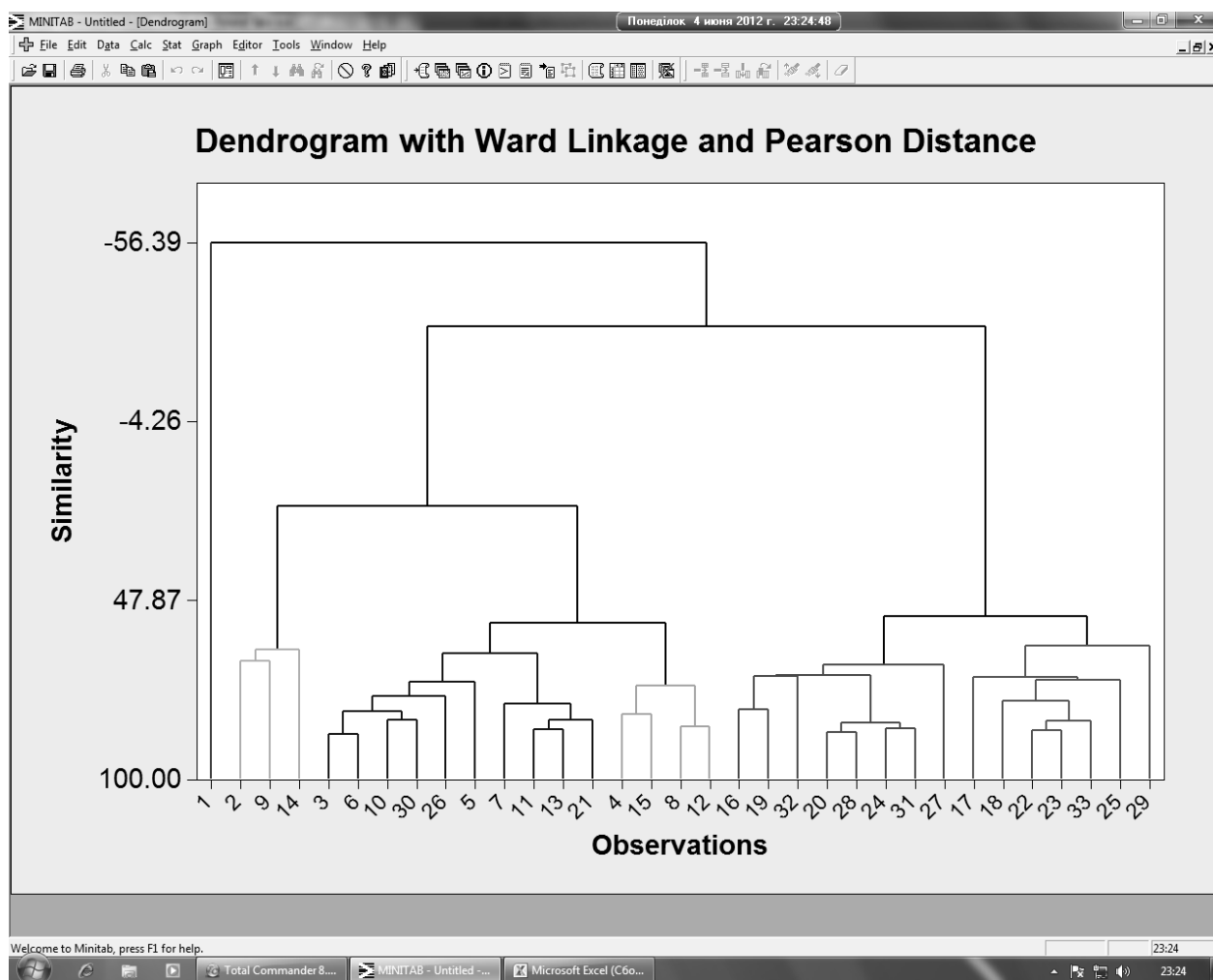


Рис. 11. Пример дендрограммы кластерного анализа

Кластер-анализ применяется при наличии многомерных совокупностей статистических показателей, сущность его заключается в объединении их в группы (кластеры) по принципу минимального размера в многомерном пространстве. В зависимости от количества объектов кластеризация и группировки выполняются последовательно в несколько шагов таким образом, чтобы на последнем шаге в одну общую группу попали все объекты. На первых шагах классификации формируются наиболее однородные группы по объектам, имеющим наибольшее сходство. Постепенно, «ослабляя» критерий относительно сходства объектов, объединяется все большее количество объектов. С каждым шагом к кластерам высшего порядка включаются целые группы административных районов, которые все сильнее различаются между собой. На последнем шаге все объекты объединяются в один кластер. В конце процедуры получают группы неоднородные. В результате кластерного анализа получают многоуровневую иерархическую классификацию, которая отражает наиболее существенные особенности взаимоотношения между объектами. Таким образом, полученные кластеры – это группа территориальных единиц,

имеющих сходные особенности развития. Такой анализ проводится для территориальных объектов по ряду показателей, для дальнейшего группировки районов и выявления устойчивых групп.

*Факторный анализ* – это многомерный статистический метод, используемый для изучения взаимосвязей между значениями переменных. Он позволяет определить группы факторов влияния и исследовать значение каждой переменной в действии фактора в разное время.

В ходе практической работы выполняется факторный анализ методом главных компонент, в MINITAB 14 путем «Stat» → «Multivariate» → «Principal Components». Факторный анализ позволяет оценить факторы развития.

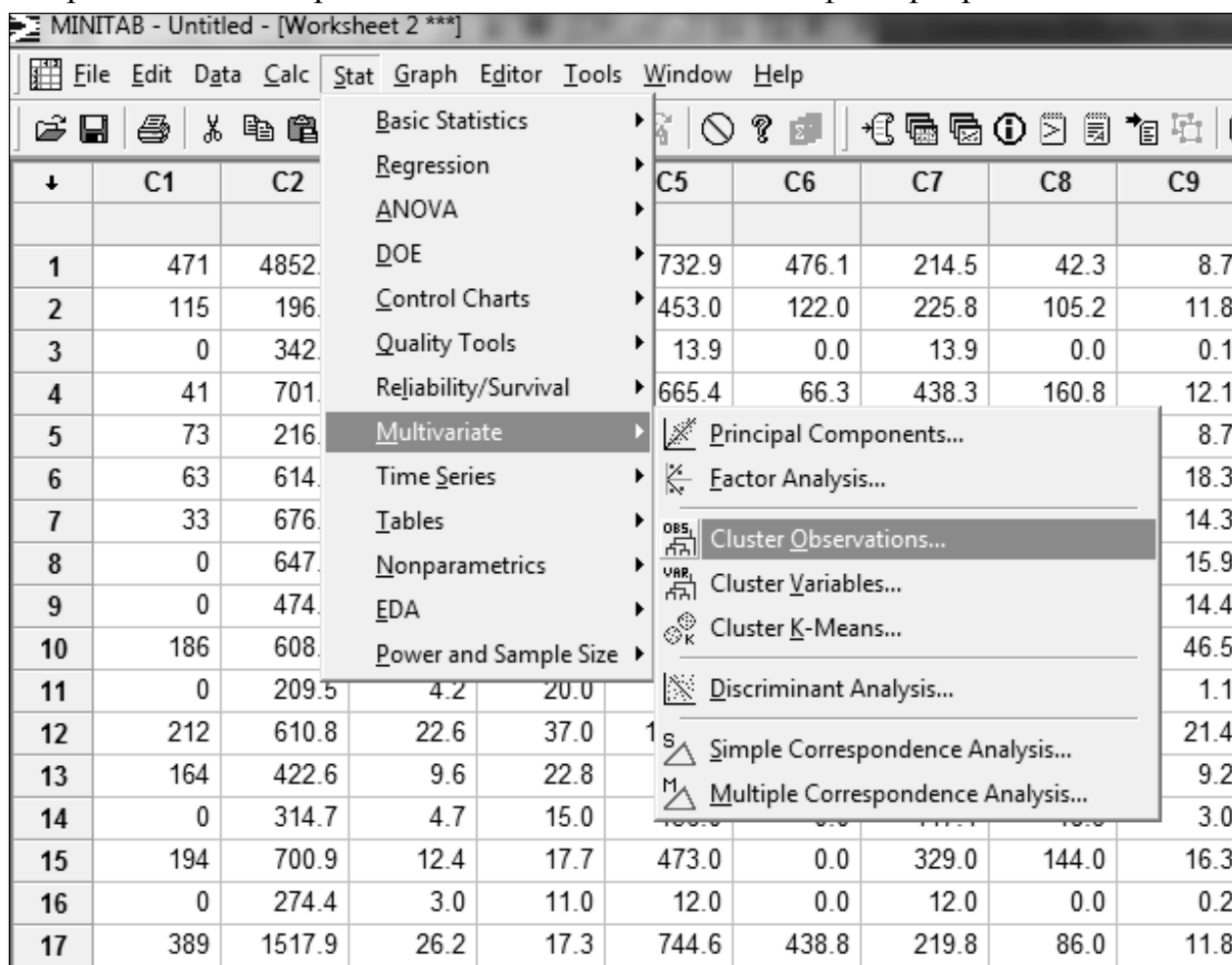


Рис. 12. Алгоритм выбора факторного анализа в программе MINITAB 14

В основе построения моделей факторного анализа лежит утверждение о том, что множество взаимосвязанных показателей, характеризующих определенный процесс, можно представить меньшим количеством гипотетических переменных – факторов и множеством независимых остатков.



Информационной базой расчетов служат пространственные ряды показателей демографического развития в разрезе административно-территориальных единиц региона за базовые годы. Набор показателей является субъективным; основным принципом выбора показателей были достоверность и сопоставимость. После избрания оптимального количества факторов, выполняется интерпретация результатов, учитывая показатели, которые сформировали выявленные факторы.

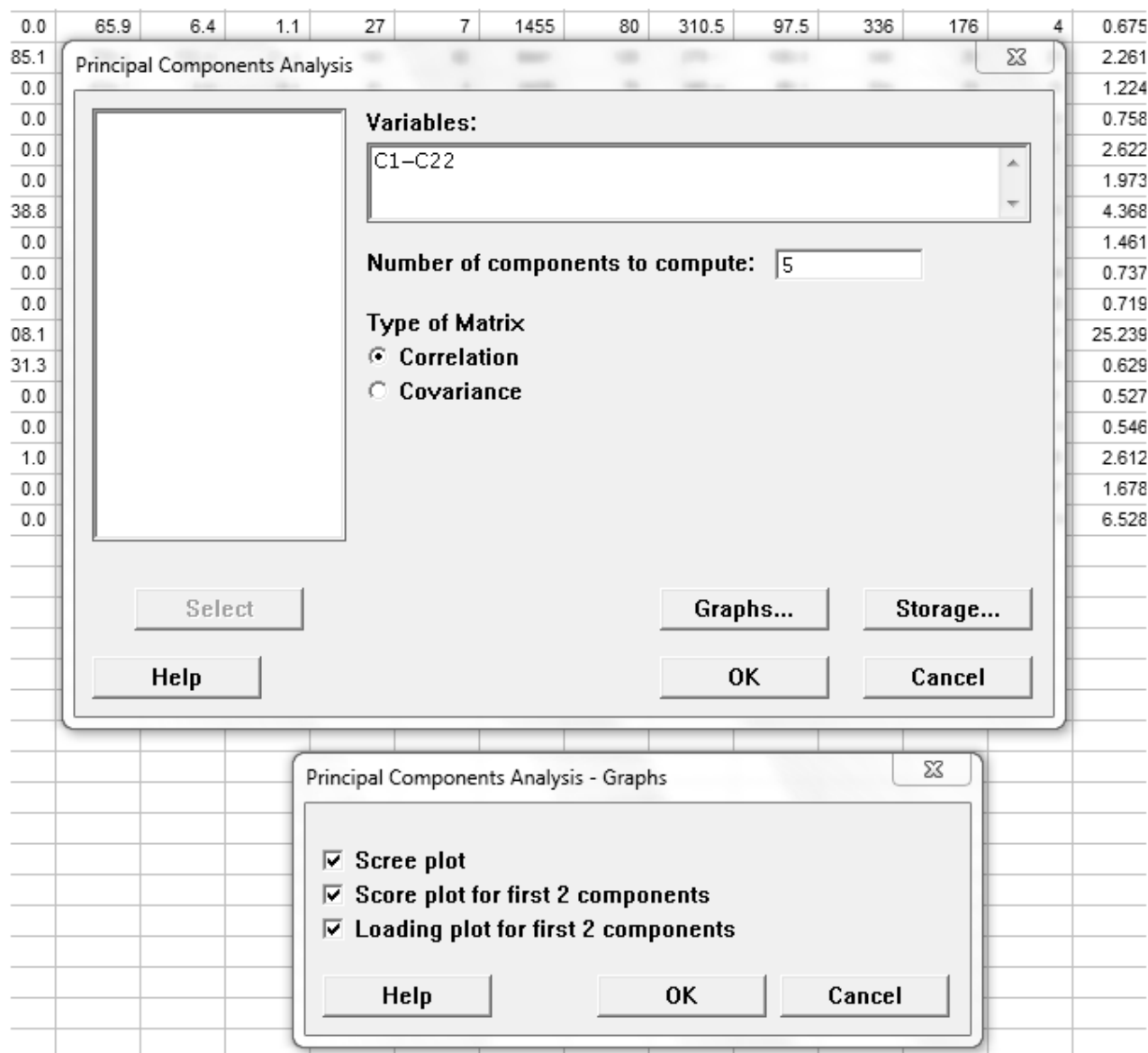


Рис. 13. Выбор исходных данных и элементов выполнения факторного анализа

Обычно факторы, влияющие на развитие общественно-пространственных процессов, характеризуются не одним, а множеством показателей, имеющих между собой определенную связь (коррелируют). Поэтому возникает потребность в их объединении в определенное количество групп по сходству

влияния, в результате чего определяются на первый взгляд “скрытые” (латентные) факторы, количество которых, разумеется, меньше количества исходных показателей. То есть факторы идентифицируются.

Применение факторного анализа предусматривает:

- во-первых, сокращение количества показателей (переменных), что на языке математической статистики называется редукцией данных;
- во-вторых, определение структуры взаимосвязей между показателями (переменными), т.е. классификацию показателей, интерпретацию факторов.

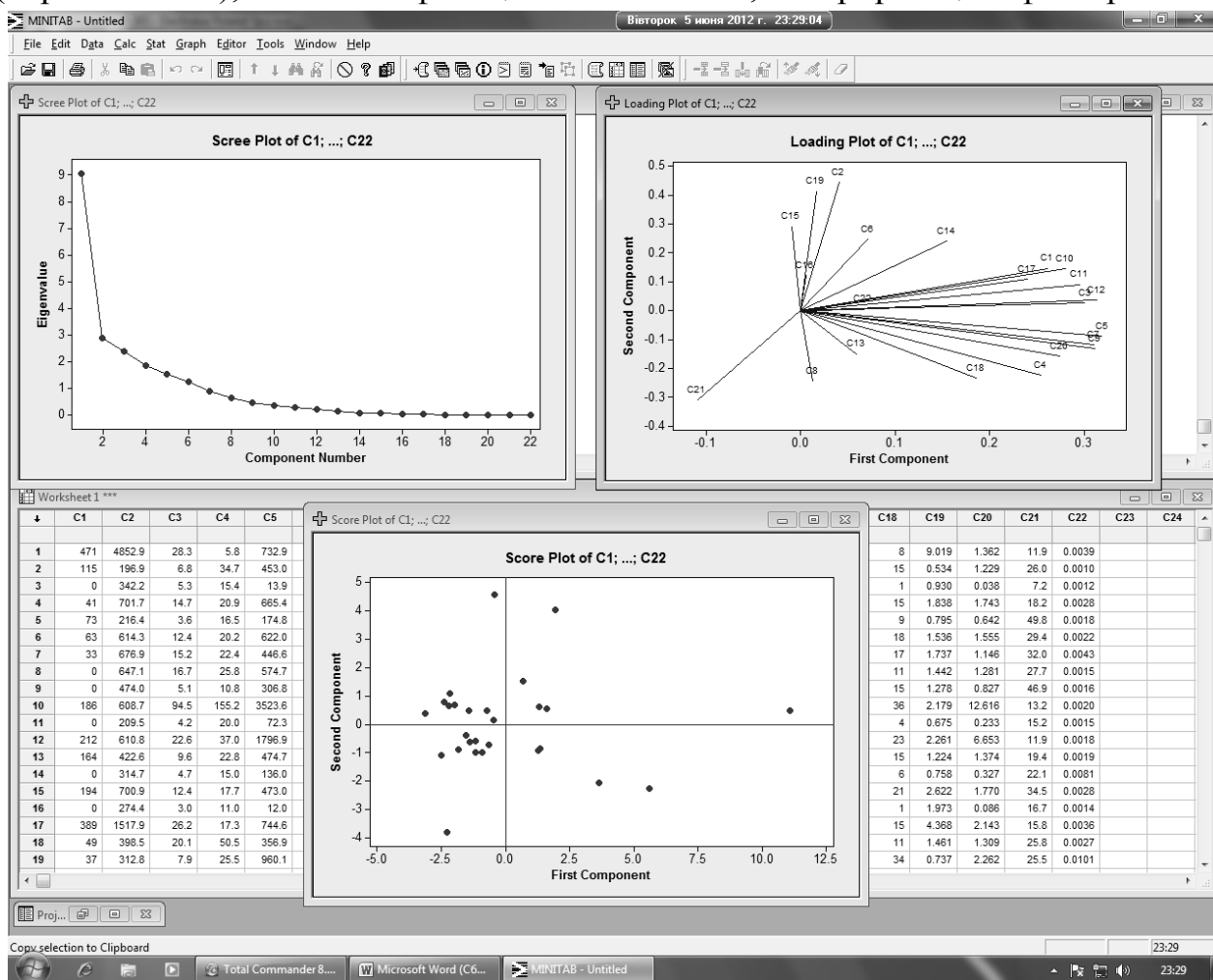


Рис. 14. Пример результата факторного анализа

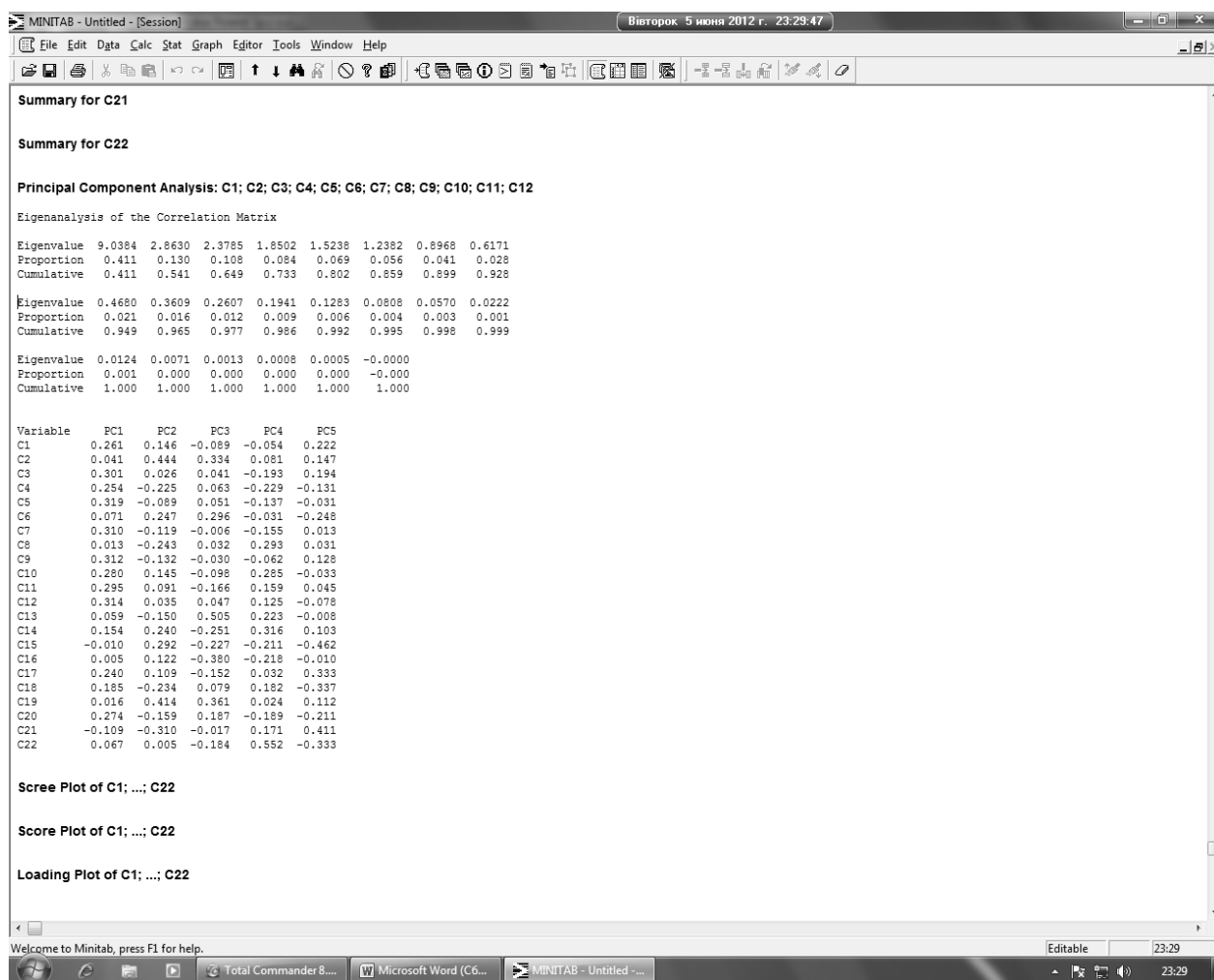


Рис. 15. Пример протокола факторного анализа

Таким образом, факторный анализ используется и как метод уменьшения количества данных, и как метод классификации одновременно. В основе построения моделей факторного анализа лежит утверждение о том, что множество взаимосвязанных показателей, характеризующих определенный процесс, можно представить меньшим количеством гипотетических переменных – факторов и множеством независимых остатков. Содержание факторного анализа заключается в линейном преобразовании  $n$ -мерного пространства в  $k$ -мерный. Другими словами, с помощью факторного анализа систему  $n$  показателей можно заменить значительно меньшим количеством ( $k$ ) факторов.

Факторные нагрузки находятся в пределах от -1 до +1. Знак "+" или "-" указывает на наличие прямой или обратной зависимости между показателем и фактором. Содержание факторов определяют показатели (переменные), имеющие наибольшие факторные нагрузки (близкие по модулю к единице).

Процедура факторного анализа включает следующие составляющие:

- Определение количества факторов. На первом шаге расчетов количество факторов избрано равным количеству показателей и рассчитаны абсолютные, относительные и кумулятивные значения дисперсии каждого из факторов. Определяют, какое количество факторов является оптимальным. Обычно пользуются одним из трех критериев:

- По критерию Кайзера (Kaiser, 1960): выбираются только факторы с дисперсией больше 1. Учитывается то, что, если фактор не выделяет дисперсию, эквивалентную хотя бы дисперсии одного показателя, то он отбрасывается. В приведенном примере следует выделять два фактора;

- По кумулятивному проценту: в качестве определяющих выбираются те факторы, которые суммарно охватывают примерно три четверти исходной информации (кумулятивный процент должен превышать 75%). В приведенном примере два первых фактора объясняют 90,5% общей дисперсии – выделяем два фактора;

- По критерию "каменистой осыпи" Кеттелла (Cattell, 1966): на графике дисперсий (plot of eigenvalue, scree plot) находится такое место, где уменьшение дисперсии слева направо максимально замедляется. Предполагается, что справа от этой точки находится только "факториальная осыпь" (термин "осыпь" заимствован из геологии, где означает обломки горных пород, которые накапливаются в нижней части скалистого склона). Согласно этому критерию в приведённом примере выделяем 2 или 3 фактора: обращение осей координат (factor rotation), идентификация и интерпретация факторов. В результате обращения получается окончательная матрица факторных нагрузок:

- Расчет факторных весов. Факторные веса (factor scores) – это показатели, которые играют роль оценок вкладов территориальных единиц в каждый из факторов. Матрица факторных весов вычисляется путем умножения матрицы исходных данных на матрицу факторных нагрузок. Они трактуются как относительные оценки проявления определенного фактора и служат основой для их группировки;

- Анализ тенденций.

Последним заданием практической работы является интерпретация полученных (в ходе выполнения всех видов анализа) результатов.

**ВОПРОСЫ К ИТОГОВОМУ МОДУЛЬНОМУ КОНТРОЛЮ**

1. Объект и цель курса.
2. Определение математических методов и моделирования.
3. Понятие «модель». Классификации моделей: натурные, аналоговые, математические.
4. История развития и использования моделей в естествознании и общественных науках.
5. Три уровня математизации общественной географии: уровень параметризации географических явлений и объектов, уровень эмпирических моделей, уровень теоретических моделей.
6. Место методов идеализации, формализации, математических методов, методов моделирования в методологии общественно-географического исследования.
7. Противоречия системного и синергетического подходов.
8. Перспективы развития моделирования в общественной географии.
9. Специфика возникновения, функционирования и развития природных и социальных объектов.
10. Многофакторность общественно-географических процессов.
11. Сложность интерпретации причинно-следственных связей в социogeосистемах.
12. Организация географической среды. Локальные и интегральные процессы.
13. Понятие о вероятности.
14. Принципиально случайный характер общественно-географических процессов.
15. Особенности и методы получения общественно-географической информации.
16. Одномерные статистические модели.
17. Понятие о случайной величине, дискретные и непрерывные случайные величины.
18. Числовые характеристики случайных величин: частота, функция распределения.
19. Оценки центра распределения: математическое ожидание, медиана, мода.
20. Оценки разброса случайной величины центральные моменты второго, третьего и четвертого порядков.
21. Основные законы одномерного распределения.
22. Выборочный метод, требования к выборочным совокупностям.
23. Понятие о надежной вероятности.
24. Точечная и интервальная оценка статистик.

25. Требования к оценкам параметров общественно-географических объектов.
26. Понятие о проверке статистических гипотез.
27. Условия применения одномерных статистических моделей.
28. Двумерные статистические модели.
29. Понятие о двумерной случайной величине.
30. Условия формирования двумерных случайных величин в общественно-географических исследованиях.
31. Особенности формирования выборочных совокупностей.
32. Особенности интерпретации связей компонентов двумерных случайных величин.
33. Понятие о форме зависимости и силе связи двух случайных величин.
34. Функциональные и статистические зависимости.
35. Двумерный регрессионный анализ, понятие об условном распределении, линейные, нелинейные уравнения регрессии, вычисления регрессионных коэффициентов.
36. Построение доверительного интервала для уравнений регрессии.
37. Корреляционный анализ, понятие о коэффициенте корреляции, корреляционное отношение.
38. Условия корректности корреляционного анализа.
39. Нелинейные преобразования случайных величин для приведения к нормальному закону распределения.
40. Особенности использования регрессионно-корреляционного анализа в общественно-географических исследованиях.
41. Многомерные статистические модели.
42. Понятие многомерной случайной величины.
43. Понятие о матрице исходных данных.
44. Многомерный корреляционный анализ, парный, частичный, множественный коэффициент корреляции. Особенности использования многомерного корреляционного анализа.
45. Многомерный регрессионный анализ, определение значимости частичных коэффициентов регрессии. Особенности применения многомерного регрессионного анализа.
46. Понятие многомерного признакового пространства, измерения расстояния в многомерном пространстве.
47. Использование теории графов для классификации общественно-географических объектов.
48. Кластер-анализ, определение дистанционных коэффициентов, принципы образования кластеров.

49. Задачи распознавания образов, понятие о собственной области объекта, решающую функцию, решающее правило. Алгоритм распознавания образов.
50. Дискриминантный анализ в задачах классификации.
51. Факторный анализ, его разновидности, интерпретация результатов.
52. Понятие о пространственной переменной, признаки и свойства пространственных переменных.
53. Различия моделей пространственных переменных от статистических моделей.
54. Общественно-географические объекты как генераторы полей пространственных переменных.
55. Изменчивость и анизотропность полей пространственных переменных.
56. Закономерная и случайная составляющие наблюдаемой изменчивости.
57. Понятие фона и аномалии.
58. Понятие о тренде.
59. Тренд-анализ как метод разделения фона и аномалии.
60. Задачи тренд-анализа в общественной географии.
61. Общая характеристика методов сглаживания случайных полей.
62. Метод скользящего статистического окна.
63. Принципы расчета локального среднего.
64. Размер скользящего статистического окна как параметр генерализации карты аппроксимации случайного поля.
65. Анализ остатков и выделение аномалий.
66. Интерпретация результатов тренд-анализа.
67. Обзор методов аппроксимации поверхности тренда единственной функцией координат пространства.
68. Принципы оценки точности аппроксимации поверхности тренда.
69. Модели алгебраических полиномов, выбор порядка полинома.
70. Модели тригонометрических полиномов.
71. Использование дифференциальных уравнений для аппроксимации поверхности тренда.

**ПРИМЕР ИТОГОВОГО МОДУЛЬНОГО КОНТРОЛЯ****I Уровень – Репродуктивно-фактажный (максимум 6 баллов)**

**Выберите один правильный ответ, за каждый правильный ответ – 1 балл)**

1. Тверждения, относительно понятия «математические методы».

- а) использование аппарата математической или формальной логики для решения конкретных задач, например, для количественного анализа данных;*
- б) некоторый материал или мысленно представляемый объект или явление, который является упрощенной версией моделируемого объекта или явления (прототипа) и в достаточной степени повторяет свойства, существенные для целей конкретного моделирования;*
- в) методы исследования явлений и процессов, основанные на замене конкретного объекта исследований (оригинала) другим, подобным ему;*
- г) научные методы описания и изучения массовых явлений, допускающих количественное выражение.*

2. Какое значение из следующих может принимать вероятность (p)?

- а)  $p = -0,5$*
- б)  $p = 0,6$*
- в)  $p = 1,5$*
- г)  $p = 3,2$ .*

3. Метод, основанный на потере информации, происходит в результате разделения объектов на кластеры и измеряемая с помощью общей суммы квадратов отклонений каждого объекта от центра своего кластера – это:

- а) «близколежащий сосед»*
- б) Варда*
- в) связь между группами*
- г) медианная кластеризация.*

4. К плохо организованной системе можно отнести:

- а) общество*
- б) приборы*
- в) физические процессы*
- г) химические процессы*

5. По какому подходу вероятность не является результатом нашего несовершенства, а является объективным атрибутом материи?

- а) гносеологическим*
- б) географическим*
- в) эпистемологическим*
- г) онтологическим.*

6. Классификация процессов, их анализ и оценка информации осуществляется на этапе математического моделирования, который называется ...

- а) содержательный анализ*
- б) количественный анализ*
- в) наблюдение-описание-измерение*
- г) прогноз-практика.*



**II Уровень – Понятийный (максимум 10 баллов)**

*Дайте определение понятию, за каждый правильный ответ – 2 балла*

Статистические методы – это

---

---

---

Частота – это

---

---

---

Мода – это

---

---

---

Дисперсия – это

---

---

---

Регрессионный анализ – это

---

---

---

**III Уровень – теоретический (максимум 24 балла)**

*Дайте развернутый ответ на вопрос:*

1. Природные и социальные объекты: специфика возникновения, функционирования и развития (8 баллов).
2. Линейный дискриминантный анализ и его применение в общественной географии (8 баллов).
3. Кластер-анализ и его модификации (8 баллов).

## ОЦЕНКА РЕЗУЛЬТАТОВ УЧЕБНЫХ ДОСТИЖЕНИЙ СТУДЕНТОВ

	Модуль 1		Модуль 2		Экзамен	Сумма баллов
	Текущий модульный контроль	Практическая работа	Текущий модульный контроль	Практическая работа		
Распределение баллов	20	10	20	10	40	100
Сумма баллов по модулю	30		30			

## ШКАЛА ОЦЕНИВАНИЯ

Сумма баллов за все виды учебной деятельности в течение семестра	Оценка ECTS	Оценка по национальной шкале	
		для экзамена, курсовой работы (проекта), практики	для зачета
90 – 100	<b>A</b>	отлично	зачислено
80-89	<b>B</b>	хорошо	
70-79	<b>C</b>		
60-69	<b>D</b>	удовлетворительно	
50-59	<b>E</b>		
1-49	<b>FX</b>	неудовлетворительно	не зачислено

## КРАТКИЙ ГЛОССАРИЙ

### **Абсолютные статистические величины**

Первоначальный вид статистических показателей, характеризующих численность совокупности, размер (объем, уровень) изучаемого явления и его факторов в конкретных границах времени и места. Абсолютные величины всегда называются и могут быть выражены в определенных единицах измерения – натуральных (км, кг, шт.), условных (усл. топливо), денежных (грн., \$), комплексных (кВт/час., Т/км), временных (год, с.) и др.

### **Аналитическое выравнивание (сглаживание)**

Эффективный метод выявления и количественной оценки тенденции изменения случайной величины в ряду наблюдений. При этом фактические уровни динамического ряда заменяют теоретическими, рассчитанными на основе уравнения регрессии. На основе фактических данных ряда динамики подбирается математическая функция – «трендовая кривая» (линейная, параболическая, показательная, экспоненциальная и др.), с помощью которой описывается основная тенденция. Время рассматривается как независимая переменная, а уровни ряда выступают как функция этой переменной  $f(t)$ .

### **Анизотропия поля**

Проявляется в зависимости значения параметра поля от направления, в котором определяется. Коэффициент анизотропии является отношением наибольшего и наименьшего значения параметра в одной точке, но определенного в разных направлениях.

### **Аппроксимация эмпирических зависимостей, полей**

Приближенное представление эмпирической зависимости, как функции времени или пространства, определенной математической функцией или уравнением регрессии (см. аналитическое выравнивание). Поля признаков объектов аппроксимируются поверхностями как единственными функциями координат пространства или построенными по принципу локальных средних.

<b>Асимметричное распределение</b>	Распределение признаков, в котором частоты по обе стороны от центра изменяются неодинаково, т.е. вершина распределения смещена. Существует правосторонняя ( $x_{cp} > M_e > M_o$ ) и левосторонняя ( $x_{cp} < M_e < M_o$ ) асимметрии. Асимметрия возникает как результат ограниченной вариации признаков в одном направлении или влияния доминирующей причины, которая приводит к смещению центра распределения.
<b>Атрибутивный ряд</b>	Последовательность статистических показателей, которую образуют по качественному признаку. Варианты располагают в их логической последовательности.
<b>Многомерная статистическая модель</b>	Модель, содержащая более двух случайных величин. Используется для установки характеристик взаимосвязи между несколькими признаками объекта исследования.
<b>Многомерный анализ</b>	Совокупность различных методов, предназначенных для изучения явлений (процессов, объектов) в многомерном признаковом пространстве.
<b>Многомерное признаковое пространство</b>	Виртуальное пространство, в котором базисом являются не обычные линейные характеристики объектов (ширина, длина, высота), а количественные показатели (параметры) признаков объектов. Измеряемость пространства не ограничивается. Положение каждого объекта в многомерном пространстве фиксируется точкой с координатами, которые соответствуют параметрам объекта.
<b>Вариационный ряд</b>	Последовательность любых чисел, расположенных в порядке возрастания (уменьшения) их величин. Вариационные ряды строятся на основе количественного группируемого признака, состоят из вариантов (отдельных значений вариационного признака, которые принимает этот признак в ряде распределения) и частот (численностей отдельных вариантов или каждой из групп вариационного ряда).

<b>Вариация</b>	Изменение значений конкретного признака при переходе от одной единицы совокупности к другой, это отдельное значение признака каждого элемента (единицы) совокупности.
<b>Выборочная совокупность (выборка)</b>	Совокупность единиц наблюдения, отобранных из генеральной совокупности по определенным правилам, которые обеспечивают репрезентативность выборки. Выборка должна быть статистически устойчивой и репрезентативной. Методы обработки выборочных данных зависят от ее объема. Существует статистика малых (до 10 элементов), средних (10 – 60 элементов) и крупных (более 60) выборочных совокупностей. Надежность результатов исследования выборочных данных возрастает с увеличением объема выборки.
<b>Выборочный метод</b>	Основной метод математической статистики, который заключается в формировании выборочной совокупности путем отбора части элементов целого (генеральной совокупности), которым присущи свойства всего сообщества. Находя статистические характеристики выборочной совокупности, их обобщают по определенным правилам на всю генеральную совокупность. Выборка репрезентативна (представительна) тогда, когда структура исследуемых признаков объекта в данной выборке незначительно отличается от структуры признаков в генеральной совокупности.
<b>Случайная величина</b>	Величина, которая принимает в результате опыта одно из множества значений, при этом появление того или иного значения этой величины к его определению (измерения) принципиально невозможно точно предсказать. Случайные величины, которые являются результатом измерения и могут принимать бесконечное число значений при последовательном определении называются непрерывными. Дискретные случайные величины имеют ограниченное фиксированное число возможных значений (подбрасывание монеты и т.д.).

<b>Относительные статистические величины</b>	Производные показатели, характеризующие количественное соотношение статистических данных. Их вычисляют как частное от деления двух абсолютных величин. Числителем есть сравниваемая (текущая) изучаемая величина, а знаменателем – база (основа) сравнения, величина с которой сравнивают и которая является своеобразным измерителем, эталоном, оптимальным уровнем. Сопоставлять можно одномерные показатели, относящиеся к разным периодам, различным объектам, территориям, а также разноизмеряемыми.
<b>Генеральная совокупность</b>	Бесконечное или конечное (достаточно большое) число элементов или компонентов, состоящих из качественно однородных показателей, это абсолютно все возможные результаты определенного события, полная совокупность точек, замеров и т.д., которые можно получить на данном объекте.
<b>Геоинформационный метод</b>	Метод исследования, который приобретает все большее значение и распространение. Заключается в создании геоинформационных систем (ГИС), которые являются современным средством сбора, хранения и анализа разнообразных сведений о территории. Использование ГИС-технологий существенно расширяет возможности анализа.
<b>Геоинформационные (геопространственные) данные</b>	Совокупность статистических показателей, имеют четкую пространственную привязку – набор координат, однозначно определяют местоположение исследуемого объекта в реальном физическом или виртуальном пространстве.
<b>Геоинформационные системы (ГИС)</b>	Информационная система, обеспечивающая сбор, хранение, отображение, распространение, анализ и преобразование геопространственных данных. ГИС содержит данные о географических объектах в цифровом представлении (растровые, векторные и т.д.)

и набор функций, реализующих геоинформационные технологии (визуализация, картографическое моделирование, пространственный анализ, проектирование и т.д.).

**Географическое пространство**

Совокупность отношений между географическими объектами, которые расположены на определенной территории (геотории) и развивающихся со временем.

**Геосистема**

Материальная система, в которой отношения между элементами опосредованы территорией (геоторией). Характеризуется целостностью, взаимосвязями и взаимодействием элементов, структурной устойчивостью и эмерджентностью.

**Гиперповерхность**

Обобщение понятия обычной поверхности 3-мерного пространства на случай  $n$ -мерного пространства. Обычно гиперповерхность задается одним уравнением связи между координатами. Если в евклидовом  $n$ -мерном пространстве гиперповерхность задается линейным уравнением, то она называется гиперплоскостью.

**Гистограмма**

Лестничный график для интервального вариационного ряда. Образующиеся прямоугольники пропорциональные по высоте частотам вариантов для каждого интервала. В случае неравных интервалов высота прямоугольников пропорциональна плотности распределения признака в конкретном интервале. Гистограмму можно превратить в полигон, соединив середины вершин столбиков линией.

**Двумерная статистическая модель**

Модель, содержащая две случайные величины и в зависимости от их типа может быть непрерывной, дискретной или смешанной. Используется для установки характеристик взаимосвязи между двумя признаками объекта исследования.

<b>Двумерный анализ</b>	Анализ двумерной случайной величины, который позволяет установить наличие зависимости между двумя случайными величинами, ее форму, тесноту, найти уравнение для ее описания. Основные виды анализа – корреляционный и регрессионный.
<b>Дендрограмма</b>	Графическое изображение в двумерной проекции результатов кластер-анализа, т.е. группировка объектов по их сходству.
<b>Детерминированная модель</b>	Модель, функционирование которой определено через известные однозначные отношения состояний и событий, и в которой один и тот же заданный вход будет всегда выдавать строго один результат (например, модель известной химической реакции).
<b>Детерминированные системы</b>	Системы, процессы в которых взаимосвязаны так, что можно отследить цепь причинно-следственных зависимостей в функциональном виде (строгое соответствие функции аргумента). Детерминизм тесно связан со степенью организации системы.
<b>Динамический (временной) ряд</b>	Совокупность значений статистических показателей, которые расположены в хронологическом порядке. Подается в виде таблицы или графика (на оси абсцисс откладывают шкалу времени, на ординате – шкалу уровней ряда).
<b>Дискриминантный анализ</b>	Разновидность многомерного анализа, предназначенного для классификации объектов по совокупности данных.
<b>Дисперсия</b>	Оценка степени отклонения значений случайной величины от центра распределения. Большие значения дисперсии свидетельствуют о большом разбросе значений случайной величины от центра распределения.



<b>Доверительная вероятность</b>	Вероятность накрытия доверительным интервалом истинного значения параметра. Длина доверительного интервала пропорциональна величине доверительной вероятности. Наиболее употребительные стандартные значения доверительной вероятности следующие: $P = 0.9; 0.95; 0.99; 0.999$ .
<b>Экстраполяция</b>	Нахождение неизвестных уровней (значений) за пределами имеющихся фактических данных динамического ряда по определенным аналитическим уравнениям тренда (математическими моделями).
<b>Закон распределения вероятностей</b>	Основное понятие теории вероятностей, которое для случайной величины определяет множество возможных событий с вероятностями их приобретения. Иначе, соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими вероятностями. Характеризуется функцией плотности распределения (отражает вероятность попадания эмпирической случайной величины в заданный интервал) и интегральной функцией распределения (отражает вероятность получения эмпирической случайной величиной значения больше или меньше заданного).
<b>Вероятность</b>	Числовая характеристика возможности того, что случайное событие произойдет в условиях, которые могут быть воспроизведены неограниченное количество раз.
<b>Интервальная оценка</b>	Отображается на числовой оси доверительным интервалом, в котором с заданной вероятностью находится оцениваемый параметр. Верхняя и нижняя границы доверительного интервала определяются величиной возможного отклонения выборочной оценки параметра от его истинного значения и называются соответственно максимальным и минимальным гарантированным значениям параметра с заданной вероятностью.

<b>Интерполяция</b>	Нахождение показателя (примерный расчет) в середине ряда на основе закономерности развития явления за период, который исследуется. Точность такого расчета зависит от стабильности показателей динамики – абсолютных приростов, темпов роста и т.д. Вычисление проводят на основе среднего абсолютного прироста или среднего темпа роста.
<b>Информационно-общественно-географическое пространство</b>	Пространство антропосферы, где существуют социум, хозяйство и природные системы, территориальная организация, свойства и связь (между собой и внешней средой), которые обусловлены сложным взаимодействием полей различных признаков, а единство и оптимизация осуществляется через интерференцию и взаимодействие их информационных полей.
<b>Информация</b>	<p>Существует несколько групп определений информации:</p> <ul style="list-style-type: none"> <li>- любые сведения, передаваемые данные по различным каналам;</li> <li>- любые сведения, данные, которые уменьшают степень неопределенности об окружающей среде;</li> <li>- мера неоднородности, разнообразия, упорядоченности, структурированности, организованности систем (атрибутивная концепция информации);</li> <li>- сохраненный выбор в точке бифуркации одного варианта развития из нескольких возможных (равноправных).</li> </ul>
<b>Кластерный анализ</b>	Вид многомерного анализа, задачей которого является разбиение заданной выборки объектов (ситуаций) на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.
<b>Ковариация</b>	Числовая характеристика взаимосвязи (взаимовлияния) случайных величин. Сущность ковариации заключается в том, что она отражает результат стохастического

взаимодействия географических объектов под воздействием большого количества факторов.

<b>Коэффициент асимметрии</b>	Количественная оценка асимметричности (скошенности) графика функции плотности распределения случайной величины.
<b>Коэффициент эксцесса</b>	Мера островершинности графика функции плотности распределения случайной величины.
<b>Коэффициент корреляции</b>	Определяется по результатам корреляционного анализа. Парный коэффициент корреляции характеризует тесноту линейной связи между двумя случайными величинами, множественный коэффициент корреляции характеризует тесноту зависимости одной случайной величины от множества других случайных величин. Частичный коэффициент корреляции отражает тесноту связи между двумя случайными величинами с учетом влияния множества других случайных величин. Коэффициент корреляции изменяется в пределах от -1 до 1, рост его значения по модулю свидетельствует об усилении связи, при прямой связи между случайными величинами коэффициента корреляции больше 0, при обратном – меньше 0. Коэффициент корреляции можно рассматривать как показатель тесноты связи только при выполнении следующих условий: а) зависимость между случайными величинами линейная б) распределение случайных величин соответствует нормальному закону.
<b>Корреляционный анализ</b>	Метод обработки статистических данных, с помощью которого измеряется теснота связи между случайными величинами.
<b>Кумулята и огива</b>	Кривые накопленных (кумулятивных) частот или частиц, их используют для изображения как дискретных так и интервальных рядов. Строя кумуляту, на абсциссе откладывают варианты, ординате – накопленные частоты. В случае построения огивы, которая является

зеркальным отражением кумуляты, наоборот. По этим кривым определяют, сколько единиц совокупности, или какова их доля не превышает определенного значения признака группирования (для дискретного ряда) или верхней границы соответствующего интервала (для интервального ряда).

**Математическая модель**

В общем случае система уравнений, описывающих изучаемый процесс или явление.

**Математическое ожидание**

Одна из основных числовых характеристик каждой случайной величины. Есть обобщенным понятием среднего значения совокупности чисел, когда элементы множества значений этой совокупности имеют разный «вес», цену, важность, приоритет, что характерно для случайных величин.

**Математические методы**

Обобщение использования аппарата математической или формальной логики для решения конкретных задач, например, для количественного анализа данных.

**Медиана**

Величина, находящаяся ровно в середине ряда значений случайной величины, расположенных в возрастающем или убывающем порядке. Медиана делит ряд значений признака на две равные части, по обе части от нее располагается одинаковое количество единиц совокупности: 50% единиц изучаемой совокупности будут иметь значение признака меньше, чем медиана, а 50% – больше.

**Метод текущей (скользящей) средней**

Метод исследования, при котором последовательно находят средние арифметические по определенным  $m$  уровням ряда, постепенно смещаясь на один уровень по ряду на каждом шагу. Найденное значение локальной средней касается того периода, который находится в середине интервала осреднения. Метод дает возможность сглаживать эмпирические данные, устанавливать тенденции динамического ряда.

<b>Метод аппроксимации поверхности тренда</b>	Описание тенденций пространственной изменчивости показателей исследуемых объектов в виде статистической поверхности. Для этого используются две группы методов: локальной средней (скользящее статистическое окно) и представления поверхности как единой функции координат пространства.
<b>Метод сглаживания случайных полей</b>	Заключается в применении скользящего статистического окна, в котором последовательно находятся локальные средние (чаще с учетом расстояния от опорных точек до центра окна) с постепенным смещением окна по пространственным координатам. В результате получают множество точек с локальными средними значениями поля, которые образуют регулярную сетку. Далее используются различные методы визуализации сглаживаний поверхностей.
<b>Мода</b>	Значение случайной величины, чаще всего встречается в совокупности наблюдений, т.е. имеет наибольшую вероятность (вес, частоту, долю). Модальное значение соответствует пику (наибольшему значению вероятности) графика функции плотности вероятности распределения случайной величины.
<b>Моделирование</b>	Метод исследования явлений и процессов, основанный на замене конкретного объекта исследований (оригинала) другим, подобным ему (моделью) в соответствии с определенными правилами, которые установлены в теории подобия и аналогии. Исследование модели дает новую информацию об оригинале.
<b>Модель</b>	Некоторый материальный или мысленно представляемый объект или явление, который является упрощенной версией моделируемого объекта или явления (оригинала) и в достаточной степени отражает свойства, существенные для целей конкретного моделирования (опуская несущественные свойства, в которых он может отличаться от оригинала). Это

материально реализованная или мысленно представления система, которая, отображая объект исследования (оригинал), объективно способна замещать его так, что ее изучение дает новую информацию об оригинале.

**Неоднородность  
(изменчивость) поля  
Нормальное  
распределение  
(распределение  
Гаусса)**

Заключается в том, что значение параметра поля зависит от координат точки и поэтому меняется в пространстве. Распределение случайной величины, которое отображается колоколообразной кривой (симметричная относительно *max* ординаты) плотности распределения. Такое распределение является результатом воздействия на значение признака или свойства объекта неограниченного количества независимых друг от друга факторов с небольшой силой воздействия, как это бывает в природе. Нормальное распределение полностью определено двумя параметрами – математическим ожиданием (средним арифметическим) и стандартным (средним квадратичным) отклонением. Значение признака при нормальном распределении преимущественно сосредотачиваются у центра распределения.

**Одномерная  
статистическая  
модель**

Статистическая модель, которая содержит одну случайную величину. В основе ее анализа является вычисление характеристик вариационного ряда: математического ожидания, дисперсии, стандартного отклонения, коэффициентов вариации, асимметрии, эксцесса и других статистик.

**Полигон частот**

Отображение сгруппированного или дискретного вариационного ряда в виде ломаной линии в прямоугольной системе координат. Значение признака (центры интервалов группировки) откладывается на оси абсцисс (X), а частоты (доли, плотность) – на оси ординат (Y).

<b>Прогнозирование</b>	Процесс предсказания будущего состояния предмета или явления на основе анализа истории его развития, систематическая информация о качественных и количественных характеристиках развития явления в перспективе (например, прогноз погоды и т.д.). Результатом прогнозирования является прогноз – знание о будущем и о вероятном развитии сегодняшних тенденций.
<b>Регрессионный анализ</b>	Статистический метод исследования зависимости между зависимой переменной $Y_1$ одной или несколькими независимыми переменными $X_1, X_2, \dots, X_p$ . В первом приближении применяют линейный регрессионный анализ (уравнение линии, плоскости, гиперплоскости). В более сложных случаях используются нелинейные уравнения (параболы, гиперболы, полиномы разных степеней и т.д.).
<b>Симметричное распределение</b>	Распределение, в котором частоты любых двух вариантов, равноотстоящих в обе стороны от центра, равны между собой.
<b>Синергетическая парадигма</b>	Возникла на основе синергетики – науки о взаимодействии и саморазвитии систем (Г. Хакен) в конце XX в. Предусматривает всесторонний анализ объекта исследования (междисциплинарный комплексный подход), учет как можно большего количества действующих факторов, ориентирует на нелинейное представление процессов. Синергетическая парадигма является продолжением системной парадигмы.
<b>Система</b>	Совокупность взаимосвязанных и взаимодействующих элементов (подсистем), которая имеет следующие признаки: целостность, устойчивость структуры, эмерджентность.

<b>Системный анализ</b>	<p>Большая группа методов для исследования систем и их взаимодействия. Основные направления: функционально-системный и структурно-системный анализ. Функционально-системный анализ главным образом направлен на исследование внешней адаптации системы, то есть, ее реакции на изменения внешней среды и взаимодействия с ней, другими системами равного порядка, реже – между собственными подсистемами. Главная задача структурно-системного анализа – исследование внутренней адаптации системы, прежде всего – структурных связей между ее элементами, условиями гомеостаза, структурной устойчивостью и т.д.</p>
<b>Социально-географическая система (социогeosистема)</b>	<p>Гетерогенная система, содержащая разные по уровню обобщения и иерархии социальные элементы или подсистемы, а также техногенные абиотические и биогенные элементы (подсистемы), находящихся во взаимодействии через потоки вещества, энергии и информации в географическом пространственно-временном континууме.</p>
<b>Социально-географический процесс (социогeопрoцесс)</b>	<p>Последовательная закономерная смена ситуаций в развитии различных социумов в историческом и географическом контексте, как изменение социальных составляющих социогeosистем в пространственно-временном континууме.</p>
<b>Статистика</b>	<p>Понятие многозначное, в научной и практической областях под ним понимают:</p> <ul style="list-style-type: none"> <li>- вид информации, то есть собранные статистические данные (итоговые цифровые показатели) о различных массовых явлениях общественной жизни;</li> <li>- форму практической деятельности, направленную на сбор, упорядочение, обработку и интерпретацию данных (системы статистических учреждений, исследовательские центры и т.д.);</li> <li>- общественную науку – специальную научную</li> </ul>



дисциплину, которая занимается разработкой теоретических положений и методов, используемых статистической практикой, а также учебную дисциплину, которую преподают в высших учебных заведениях всех уровней для подготовки специалистов высшей квалификации;

- «меру», «оценку» или «характеристику», то есть определенные статистические показатели, например, обобщенные показатели выборочной совокупности (средняя, доля) и т.д.

**Статистическая  
закономерность**

Количественное представление закономерностей изменений в пространстве и времени массовых явлений и процессов общественной жизни, состоящих из множества элементов (единиц совокупности). Это форма проявления причинной связи, выраженная в последовательности, регулярности, повторяемости событий, если причины, вызвавшие эти события не изменяются или имеют незначительные изменения.

**Статистическая  
совокупность**

Множество значений параметра исследуемого объекта (однородных в определенном отношении), соединенных общими свойствами, условиями и причинами существования и развития.

**Статистический  
показатель**

Количественная или полуколичественная характеристика свойства (признака) объекта исследований, которая определенным образом отражает механизмы процессов или состояние объекта. Количественно определенные показатели называются параметрами. Статистические показатели делятся на абсолютные, относительные и средние величины и образуют статистические совокупности.

**Статистический ряд  
распределения**

Упорядоченное распределение единиц изучаемой совокупности на группы по группированному (вариативному) признаку. Они характеризуют состав (структуру) изучаемого явления, позволяют судить об однородности совокупности, предела ее изменения,

закономерностях развития изучаемого объекта. В зависимости от признака статистические ряды распределения делятся на: атрибутивные (качественные) и вариационные (количественные), последние делятся на дискретные и интервальные.

**Статистические  
методы**

Набор методов анализа статистических данных, например, в социально-экономической географии о хозяйстве и населении регионов и стран для выяснения особенностей пространственного взаимодействия различных территориально-экономических систем. В этом аспекте важную роль играет статистика как эффективный инструмент исследования экономического и социального положения конкретной территории.

**Стохастическая  
(статистическая)  
связь**

Вид связи между случайными переменными, для которого характерно, что каждому значению факторного (независимого) признака соответствует определенное множество (спектр) значений результативного (зависимого) признака. То есть причинная зависимость проявляется не в каждом отдельном случае, а в общем, при большом количестве наблюдений. В отличие от функциональных, стохастические связи неоднозначны и характеризуются вероятностью. Чаще всего такие связи оказываются как согласованность вариации двух или более признаков.

**Общественно-  
географический  
процесс**

Последовательная и закономерная смена общественно-географических явлений. Является одним из основных понятий социально-экономической географии и обобщает все процессы, протекающие в обществе и техносфере (антропосфере).

**Точечная оценка**

Соответствует точке на числовой оси и является случайной величиной. Должна соответствовать следующим требованиям: убедительность, несмещенность, максимальная эффективность. Точечная оценка не содержит информацию о точности полученного результата.

**Тренд-анализ**

В наиболее общем виде тренд-анализ является методом исследования определенных пространственно-временных закономерностей в развитии общественно-географического процесса, отраженных в наборе статистических совокупностей параметров объекта исследования. Включает несколько групп методов анализа эмпирических (статистических) данных для выделения закономерной (фоновой) и случайной (аномальной) составляющих поля параметра определенного признака объекта исследования. Согласно этому обычно любое поле параметра (которое рассматривается как совокупность случайных величин) разлагается на закономерную (трендовую), аномальную (случайную) составляющие и так называемый остаток – разницу между ними. В таком случае говорят о снятии тренда и совокупность остатка рассматривают как отдельный объект исследования.

**Факторный анализ**

Большая группа статистических методов анализа многомерных статистических совокупностей, для которых общей задачей является выявление латентных (скрытых) гипотетических факторов влияния на исследуемый процесс или объект. Как правило, эти факторы имеют определенную содержательную интерпретацию, более полно раскрывают особенности взаимодействия отдельных факторов, отраженных в статистических показателях.

**Фон та аномалия**

С точки зрения математической статистики фон является определенным средним (нормативным) уровнем значений поля любого параметра, определенным с учетом доверительного интервала. Значение поля, выходящие за границы доверительного интервала, рассматриваются как случайные отклонения и составляют аномалии. Очень часто выделения аномалий является одной из основных задач исследования и достигается применением тренд-анализа.

<b>Формализация</b>	Представление объекта исследования в виде, пригодном для анализа средствами формальной (математической) логики. Заключается в преобразовании качественных (концептуальных) представлений о структуре, свойствах, особенностях динамики объекта в формальное описание объекта в виде однозначно определенных математических зависимостей.
<b>Функциональная связь</b>	Характеризуется полным соответствием между изменением независимого признака (аргумента) и исходной величины (функции). При функциональной связи каждому возможному значению факторного признака ( $x$ ) соответствует одно и только одно четко определенное значение результативного признака ( $y$ ). Благодаря этому функциональную зависимость можно описать математическими формулами. Такая связь есть в физических, химических процессах – зависимость длины ртутного столба от температуры окружающей среды. Функциональные связи присущи преимущественно техническим системам и частично – природно-техногенным системам. В общественных процессах – это связь между элементами расчетных формул показателей – аддитивный ( $a+b+c$ ) или мультипликативный ( $a=bc$ , $c=a/b$ ), а также зависимость средних величин от структуры совокупности. Функциональные зависимости изучают точные науки – математика, физика, химия. Для исследования общественных явлений их используют, главным образом, как модели более сложных стохастических связей.
<b>Функция плотности распределения</b>	Характеризует вероятность попадания выборочного значения случайной величины в заданный интервал.
<b>Функция распределения</b>	Характеризует вероятность значения случайной величины меньше заданного.

<b>Центральные моменты</b>	Являются накопленной суммой отклонений частных значений случайной величины от центра распределения (среднего значения) в определенной степени, отнесенной к стандартному отклонению в той же степени. Используются для определения дисперсии (вторая ступень), асимметрии (третья ступень), эксцесс (четвертая ступень).
<b>Частота</b>	Число повторений отдельного значения дискретной случайной величины, или попадания значения случайной величины в определенный интервал группировки.
<b>Шкалы измерения параметров</b>	<p>В географии используются четыре шкалы количественной оценки параметров географических объектов:</p> <ol style="list-style-type: none"> <li>1. <i>Номинальная шкала</i> – используется для классификации объектов по принципу равенства их свойств, при этом число является кодом качественного показателя объекта.</li> <li>2. <i>Порядковая шкала</i> – применяется, когда объекты можно расположить в определенном порядке в зависимости от изменения любого свойства, но эти изменения нельзя оценить количественно. Это классификации, построенные по схеме высоко–средне–низко, или сильно–средне–слабо без количественных критериев выделения классов.</li> <li>3. <i>Интервальная шкала</i> – применяется, когда интервалы перехода объектов из класса в класс равные и имеют количественно определенные пределы, но отсутствует точка абсолютного нуля. Примером служат температурные шкалы Кельвина, Цельсия, Фаренгейта,</li> <li>4. <i>Шкала отношений (относительная шкала)</i> – повторяет все свойства интервальной шкалы, но имеет точку абсолютного нуля (начальную точку отсчета координат). Эта шкала является самой высшей шкалой измерений и применяется, когда параметры объектов имеют количественную оценку.</li> </ol>

Навчальне видання

Немець Костянтин Аркадійович  
Сегіда Катерина Юріївна

НАВЧАЛЬНО-МЕТОДИЧНИЙ ПОСІБНИК

з дисципліни

**«СТАТИСТИЧНІ МЕТОДИ І ОБРОБОТКА ГЕОІНФОРМАЦІЇ»**